
W O R L D

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:03:21 2000; MasPar time 8.27 Seconds
Tabular output not generated. 638.314 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVRLRLALIFAIVTTW.....KYEGWPELLEMEGCMPPKPF 223

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 33.108; Variance 140.762; scale 0.235

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1686	100.0	223	1 W80475	Protein encoded by a c	1.51e-153
2	993	58.9	135	1 W46878	Protein sequence encod	2.59e-84
3	993	58.9	135	1 W06547	Human colon specific g	2.59e-84
4	108	6.4	618	1 R97984	DmORP1 potassium chann	2.45e+00
5	102	6.0	993	1 W48720	Human mitochondrial is	6.95e+00
6	95	5.6	187	1 P81244	Sequence of yellow tai	2.28e+01
7	95	5.6	476	1 W35714	Gentian flavonoid 3'-h	2.28e+01
8	94	5.6	642	1 W62596	Human nuclear protein	2.70e+01
9	94	5.6	642	1 W80950	Amino acid sequence of	2.70e+01
10	92	5.5	146	1 R42381	Haemin-binding protein	3.77e+01
11	92	5.5	176	1 R42379	Haemin-binding protein	3.77e+01
12	92	5.5	178	1 R42374	Haemin-binding protein	3.77e+01
13	91	5.4	191	1 R10967	Bovine Somatotropin an	4.44e+01
14	91	5.4	217	1 R10955	Bovine Somatotropin an	4.44e+01
15	91	5.4	217	1 W40449	Bovine growth hormone	4.44e+01
16	91	5.4	533	1 R26153	HUG-Br1.	4.44e+01
17	91	5.4	534	1 R26154	HUG-Br2.	4.44e+01
18	89	5.3	191	1 R10958	Bovine Somatotropin an	6.17e+01
19	89	5.3	204	1 W27340	Yellow tail/fin tuna g	6.17e+01
20	89	5.3	204	1 W25120	Yellowtail tuna growth	6.17e+01
21	90	5.3	217	1 R21115	Bovine growth hormone	5.24e+01
22	90	5.3	217	1 W40457	Bovine growth hormone	5.24e+01
23	89	5.3	275	1 W77603	Staphylococcus aureus	6.17e+01

RESULT 1

ID W80475 standard; Protein; 223 AA.
AC W80475;
DT 28-JAN-1999 (first entry)
DE Protein encoded by a consensus sequence of CS141 EST.
KW EST sequence: CS141; gastrointestinal tract; cancer; Barret's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polyyps; Crohn's disease;
KW ulcerative colitis; pancreatitis.
OS Synthetic.
OS Homo sapiens.
PN W09844133-AT.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABBO) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Grados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI: 98-568280/48.
DR N-PSDB; V63516.
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
develop products for the diagnosis and treatment of e.g. cancers,
gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
pancreatitis
FS Claim 17; Pages 95-96; 116pp; English.
CC The present sequence is encoded by a consensus sequence of a set of
contiguous and partially overlapping EST sequences designated CS141.
CC The sequences are isolated from a cDNA library made from gastrointestinal
tract tumour and normal tissues. The CS141 gene is useful as a marker for
gastrointestinal tract disorders. The methods and products can be
used in detecting, diagnosing, staging, monitoring, prognosticating,
preventing, or determining the predisposition to diseases and
conditions of the gastrointestinal tract, such as gastrointestinal
tract cancers, Barret's esophagus, gastric ulcer, gastritis,
leiomyoma, polyyps, Crohn's disease, ulcerative colitis, pancreatitis.
SQ Sequence, 223 AA;

ALIGNMENTS

24	90	5.3	462	1	W69359	Pyridine nucleotide tr	5.24e+01
25	89	5.3	614	1	W57326	Mitochondrial glycerol	6.17e+01
26	89	5.3	614	1	W30682	Glycerol-3-phosphate d	6.17e+01
27	89	5.3	614	1	W60257	Klebsiella pneumoniae	6.17e+01
28	89	5.3	679	1	W96998	The gibbon ape leukemi	6.17e+01
29	89	5.3	679	1	R72969	GALV receptor protein.	6.17e+01
30	89	5.3	679	1	R10848	Human gibbon ape leuka	6.17e+01
31	89	5.3	679	1	W70498	Human sodium-lithium c	6.17e+01
32	88	5.2	166	1	W59885	Amino acid sequence of	7.27e+01
33	88	5.2	190	1	P60014	Sequence of ovine grow	7.27e+01
34	88	5.2	191	1	R35432	Ovine growth hormone.	7.27e+01
35	87	5.2	191	1	R40169	Ovine growth hormone.	8.55e+01
36	87	5.2	194	1	R03253	Recombinant Human Soma	8.55e+01
37	88	5.2	217	1	P40015	Ovine pre-growth hormo	7.27e+01
38	88	5.2	217	1	R05794	Ovine growth hormone-1	7.27e+01
39	88	5.2	217	1	P90914	Goat growth hormone pr	7.27e+01
40	88	5.2	418	1	W98535	H. pylori ORF 1479 pr	7.27e+01
41	87	5.2	418	1	W55528	H. pylori GRP 29ge3032	8.55e+01
42	87	5.2	713	1	R10052	Cyclomaltodextrin gluc	8.55e+01
43	87	5.2	713	1	R06110	Sequence of cyclomalto	8.55e+01
44	88	5.2	1835	1	R92316	Peripheral nervous sys	7.27e+01
45	86	5.1	1029	1	R38862	GC-A.	1.00e+02

QY 61 PANYFAKICSGAANVVTGTCFEDRMIMSPVKNVGRGLNLTALVNGTTGAVLGOKAFDM 120
 Db 121 YSGDVHMLVKFLKEIPGGALVLVASYDDPGTKMNDSEKRLKFLSDLGSSYAKQLGFRDSWVF 180
 QY 121 YSGDVHMLVKFLKEIPGGALVLVASYDDPGTKMNDSEKRLKFLSDLGSSYAKQLGFRDSWVF 180
 Db 181 IGAKDLRGKSPFEQFLKNSPDNTNKYEGWPELLEMEGCMPPKPF 223
 QY 181 IGAKDLRGKSPFEQFLKNSPDNTNKYEGWPELLEMEGCMPPKPF 223

RESULT 2

ID W46878 standard; Protein; 135 AA.
 AC W46878;
 DT 22-JUN-1998 (first entry)
 DE Protein sequence encoded by a colon-specific gene.
 KW Colon-specific gene; probe: detection; expression; human;
 KW diagnostic assay; colon cancer; antibody; screening.
 OS Homo sapiens.
 PN US5733748-A.
 PD 31-MAR-1998.
 PF 06-JUN-1995; 469667.
 PR 06-JUN-1995; US-469667.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen C, Yu G;
 DR WPI: 98-729823/20.
 DR N-PSDB; V16671.
 PT Colon-specific nucleic acids - useful as probes for detecting colon
 PT cancer micrometastases
 PS Claim 1; Fig 4; 51pp; English.
 CC W46876-80 and W4682-85 represent proteins encoded by colon-specific
 CC genes. The polynucleotides encoding these proteins can be used
 CC as probes to detect expression of the corresponding human genes,
 CC e.g. in diagnostic assays for detecting micrometastases of colon cancer.
 CC Recombinant cells containing the polynucleotides can be used to
 CC produce the proteins, in order that antibodies can be raised and
 CC used in further screening or diagnostics.
 SQ Sequence 135 AA;

Query Match 58.9%; Score 993; DB 1; Length 135;
 Best Local Similarity 100.0%; Pred. No. 2.59e-84;
 Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 60
 QY 89 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 148
 Db 61 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 120
 QY 149 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 208
 Db 121 PELLEMEGCMPPKPF 135
 QY 209 PELLEMEGCMPPKPF 223

RESULT 3

ID W06547 standard; Protein; 135 AA.
 AC W06547;
 DT 13-MAR-1997 (first entry)
 DE Human colon specific gene CSG4 polypeptide fragment.
 KW Colon specific gene; CSG4; colon cancer; metastasis; diagnosis;
 KW therapy; antibody; vaccine.
 OS Homo sapiens.
 PN W09639419-A1.
 PD 12-DEC-1996.
 PF 06-JUN-1995; U07289.
 PR 06-JUN-1995; WO-007289.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Yu G;
 DR WPI: 97-043054/04.
 DR N-PSDB; T45883.
 PT Human colon specific genes and their expression products - detection

PT of which, in non-colon tissue samples, can be used as indication of
 PT colon cancer metastasis
 PS Claim 8; Fig 4; 60pp; English.
 CC Novel polypeptides (W06545-53) are encoded by cDNA clones (see also
 CC T45880-92) corresponding to 13 human colon specific genes,
 CC designated CSG1, CSG2, etc., that are primarily expressed in
 CC tissues derived from the colon. Recombinant CSG polypeptides can
 CC be produced in transformed host cells. They are useful diagnostic
 CC markers for colon cancer and for colon cancer metastasis and can
 CC also be used to screen for (ant)agonist cpds. of therapeutic or
 CC diagnostic value. Antibodies raised against the colon-specific
 CC polypeptides may be used to target colon cancer cells or as part
 CC of a colon cancer vaccine.
 SQ Sequence 135 AA;

Query Match 58.9%; Score 993; DB 1; Length 135;
 Best Local Similarity 100.0%; Pred. No. 2.59e-84;
 Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 60
 QY 89 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 148
 Db 61 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 120
 QY 149 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 208
 Db 121 PELLEMEGCMPPKPF 135
 QY 209 PELLEMEGCMPPKPF 223

RESULT 4

ID R97984 standard; Protein; 618 AA.
 AC R97984;
 DT 16-OCT-1996 (first entry)
 DE DmORF1 potassium channel protein.
 KW transmembrane helix; N-glycosylation site; potassium-agonist;
 KW potassium-antagonist; drug screening; insecticide; cardiac disorder.
 OS Drosophila melanogaster.
 FH Key Location/Qualifiers
 FT domain 6..27 /note= "Transmembrane domain M1"
 FT modified_site 58..60 /note= "N-glycosylation site"
 FT domain 95..111 /note= "Pore-forming domain H5-1"
 FT domain 120..140 /note= "Transmembrane domain M2"
 FT domain 171..191 /note= "Transmembrane domain M3"
 FT domain 208..224 /note= "Pore-forming domain H5-2"
 FT domain 242..263 /note= "Transmembrane domain M4"
 FT WO9613520-A1.
 PN 09-MAY-1996.
 PD 25-OCT-1995; U14364.
 PF 31-OCT-1994; US-332312.
 PA (AMCY) AMERICAN CYANAMID CO.
 PI Pausch MH, Price LA;
 DR WPI: 96-239450/24.
 DR N-PSDB; T33201.
 PT Potassium channel genes from Drosophila melanogaster and
 PT Caenorhabditis elegans - useful in assaying substances to determine
 PT effects on cell growth, and in inhibiting nematode and insect pests
 PS Claim 26; Page 37-39; 79pp; English.
 CC This potassium channel sequence is encoded by the DmORF1 gene from
 CC Drosophila melanogaster, and has 2 pore-forming domains between 4
 CC hydrophobic transmembrane helix domains. Each pore-forming H5
 CC domain contains a Y/F-G dipeptide motif required for potassium
 CC selectivity. The protein contains a single N-terminal

CC asparagine-linked glycosylation site. The protein is predicted to span the membrane 4 times, with N- and C-termini within the cell, and the N-glycosylation site and H2 domains on the cell exterior, allowing permeation of the membrane by the pore-forming domains from the outside, a requirement for functional potassium channel formation. The protein may be expressed in a heterologous host cell to assay substances to determine effects on cell growth. CC Potassium-agonists or potassium-antagonists identified by this method may be used as insecticides or in therapy of cardiac disorders, etc CC Sequence 618 AA;

Query Match 6.4%; Score 108; DB 1; Length 618;
Best Local Similarity 33.3%; Pred. No. 2.45e+00;
Matches 20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;

Dd 97 FFAFTVCTVGVGNISPTT-FAGRMIMAY-SVIGIPVNGILFAGI-GEVFG-RTEAYI 152
QY 64 YFAFKICSGAA-NVVGPTMCFEDRMIMSPVKNVGRGLNALVNGTTGAVLGQAFD-MY 121

RESULT 5
ID W48720 standard; Protein; 993 AA.
AC W48720;
DT 19-AUG-1998 (first entry)
DE Human mitochondrial isoleucyl-tRNA synthetase.
KW Mitochondrial isoleucyl-tRNA synthetase; T-cell; pathogen.
OS Homo sapiens.
PN US5759833-A.
PD 02-JUN-1998.
PF 27-MAY-1994; 250852.
PR 06-JUN-1995; US-468557.
PR 27-MAY-1994; US-250852.
PA (CANC-) CANCER INST JAPANESE FOUND CANCER.
PA (CUBI-) CUBIST PHARM INC.
PI Kranz JE, Schimmel PR, Shiba K;
DR WPI; 98-332142/29.
DT N-PSDB; V18326.
PR Human isoleucyl-tRNA synthetase DNA - useful for producing recombinant proteins, tester strains, etc.
PS Claim 3; Columns 47-52; 46pp; English.
CC The present sequence is that of a human mitochondrial isoleucyl-tRNA synthetase. The cDNA encoding for the mitochondrial isoleucyl-tRNA synthetase was isolated from a human T-cell cDNA library. The cDNA can be expressed using expression vectors to produce the corresponding recombinant protein. The protein can be used in assays to test substances known to inhibit the isoleucyl-tRNA synthetase or other tRNA synthetases of pathogenic organisms. Also, expression of the isoleucyl-tRNA synthetase in a tester strain can be useful for testing substances which are capable of inhibiting its activity.
SQ Sequence 993 AA;

Query Match 6.0%; Score 102; DB 1; Length 993;
Best Local Similarity 30.1%; Pred. No. 6.95e+00;
Matches 22; Conservative 23; Mismatches 23; Indels 5; Gaps 5;

Dd 887 VIBPGLIFELIEMQSESTSQ-LN-ELMMASETLLAQEPREM-TADVIELKGKFLI 943
QY 76 VVGPTMCFEDR-MIMSPVKNVGRGLNALVNGTTGAVLGQAFD-MYSGDVHML-VKFLK 133

Dd 944 NLEGGDIRRESSY 956
QY 134 EIPGGLVLVASY 146

RESULT 6
ID P81244 standard; protein; 187 AA.
AC P81244;
DT 16-NOV-1990 (first entry)
DE Sequence of yellow tail fish pre-growth hormone
KW pyGHI; somatotropin.
OS Yellow tail fish.
FH Key Location/Qualifiers

FT protein 3. 187
FT /note="185 AA SQ, see CC"
PN J63152985-A.
PD 25-JUN-1988.
PF 23-JUL-1987; 184083.
PR 24-JUL-1986; JP-174385.
PR 3-JUL-1987; JP-184083.
PA (NAKA) Nakajima K.
DR WPI; 88-216878/31.
PT Escherichia coli plasmid used for growth of yellow tail fish - comprises plasmid pS14001 derived from ampicillin resistant gene
PT obtd. from pituitary tissue of yellow tail fish
PS ; Fig 4; 7pp; Japanese.
CC The E.coli plasmid comprises plasmid pS14001 derived from ampicillin resistant gene contg. E.coli plasmid pBR322 and yellow tail pre-growth hormone having cyclic double chain plasmid pyGHI of 3851 BP, and yellow tail growth hormone of 185 and 187 (p81244) AA sequences contd. in pyGHI.
CC The plasmids were introduced into E.coli host cell DH1. The plasmids are used for cultivation of young yellow tail, by E.coli and Bacillus subtilis having recombinant plasmid for yellow tail pre-growth hormone structural gene amplification.
SQ Sequence 187 AA;

Query Match 5.6%; Score 95; DB 1; Length 187;
Best Local Similarity 29.8%; Pred. No. 2.28e+01;
Matches 14; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

Dd 69 SSVLKLSTSYRLVSEWFFSSFLSGSALRNQISPRSELKTEIQ 115
QY 4 SGVLRLLALFAIVTWTFIRMSF-SMKITRLPRLWLASPTKEIQ 49

RESULT 7
ID W35714 standard; Protein; 476 AA.
AC W35714;
DT 27-MAR-1998 (first entry)
DE Gentian flavonoid 3'-hydroxylase.
KW Flavonoid 3'-hydroxylase; pigmentation; flower colour;
KW transgenic plant; gentian.
OS Gentiana triflora Pall. var japonica Hara.
PN WO9732023-A1.
PD 04-SEP-1997.
PF 28-FEB-1997; AU0124.
PR 01-MAR-1996; AU-008386.
PA (FLOR-) FLORIGENE LTD.
PI Brugliera F, Holton TA, Michael MZ;
DR N-PSDB; T94664.
PT Novel flavonoid 3'-hydroxylase(s) from flowering plants - and corresponding DNA, used in the manipulation of pigmentation in plants
PT plants
PS Claim 23; Page 181-184; 234pp; English.
CC This polypeptide comprises gentian flavonoid 3' hydroxylase (F3'H). Its amino acid sequence was deduced from a cDNA clone (see T94664) isolated from a petal cDNA library. F3'H acts on dihydrokaempferol to produce dihydroquercetin and on naringenin to produce eriodictyol. Isolated F3'H nucleic acids (see T94655-65) can be used to provide genetic constructs capable of modulating the hydroxylation of flavonoid compounds in a plant or in the cells of a plant, e.g. by de novo expression, over-expression, suppression, antisense inhibition and ribozyme activity. Modulation of F3'H is useful for the manipulation of pigmentation (flavonoids contribute to a range of colours from yellow to red to blue) in plants, especially flowering plants such as petunia, carnation, chrysanthemum, rose, snapdragon, tobacco, cornflower, pelargonium, lisianthus, gerbera, apple, iris, lily, african violet and morning glory.
SQ Sequence 476 AA;

Query Match 5.6%; Score 95; DB 1; Length 476;
Best Local Similarity 24.5%; Pred. No. 2.28e+01;
Matches 23; Conservative 23; Mismatches 45; Indels 3; Gaps 2;

Db 88 LKIKSVHLEFSSKALDDFOVHRHEFTICILIRAIASGGHAPVNLGKLGVCVTNALARVML 147
 QY 56 LKPCPVANVFAFKICSGAANVVGPTMCFEDRMIPVKN--NVGRGLNALVNGTTGAVL 113
 Db 148 GRVFEQDGGENPHADEF-KSMVVEIWLAFN 180
 QY 114 GKAFDMYSGYDMLVHLVKFLKEIPGGALVLVASYD 147

RESULT 8

ID W62596 standard; Protein; 642 AA.
 AC W62596;
 DT 26-OCT-1998 (first entry)
 DE Human nuclear protein HEC.
 KW HEC gene; highly expressed in cancer; human; nuclear protein;
 KW mitosis; cell cycle; cell proliferation; malignancy; vaccine;
 KW antibody.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Modified_site 165
 FT 254..621
 FT Region /label= O-phosphorylated
 FT /note= "leucine heptad repeat region"
 PN W09827994-Al.
 PD 02-JUL-1998.
 PR 18-DEC-1997; U23385.
 PR 20-DEC-1996; US-033600.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Chen P, Chen Y, Lee W, Riley DJ;
 DR WPI; 98-377401/32.
 DR N-PSDB; V38564.
 PT Nucleic acid encoding human nuclear protein HEC modulating mitosis -
 PT useful to, e.g control cell malignancies and other cell growth
 PT abnormalities at mitosis stage and to produce HEC protein and
 PT peptide(s)
 PS Claim 1; Page 63-65; 93pp; English.

CC This protein comprises a novel human nuclear protein, designated
 CC HEC (highly expressed in cancer), that appears to be crucial for
 CC normal mitosis. Its amino acid sequence was deduced from an
 CC isolated B cell cDNA clone (see V38564). HEC localises to the
 CC nuclei of interphase cells and redistributes to centromeres during
 CC M phase. Ectopic expression of a mutant HEC containing only the
 CC leucine heptad repeats results in cells being unable to divide more
 CC than once. Inactivation of HEC results in disordered sister
 CC chromatid alignment and separation, and formation of non-viable
 CC cells with multiple, fragment micronuclei. HEC interacts through
 CC its leucine heptad regions with several proteins involved in
 CC mitosis, including nek2, sbl.8 and 2 different regulatory subunits
 CC of the 26S proteasome, MSL1 and p45. These properties suggest
 CC potential roles of HEC in modulating proteins important for spindle
 CC attachment to kinetochores, sister chromatid movement, and M phase
 CC progression. HEC nucleic acids and polypeptides are useful in
 CC compositions to control cell malignancies and other cell growth
 CC abnormalities at the cell mitosis stage. HEC peptides are useful
 CC in vaccines and may be administered to cells to disrupt chromatid
 CC separation and so modulate cell cycle progression (claimed). HEC
 CC specific antibodies are useful in purifying native or recombinant
 CC HEC and to detect HEC protein/peptides in samples (claimed)
 CC immunodetection kits are provided). They can also be administered
 CC to disorder sister chromatid alignment and separation in an
 CC interphase cell, so disrupting mitosis (claimed).
 SQ Sequence 642 AA;

Query Match 5.6%; Score 94; DB 1; Length 642;
 Best Local Similarity 40.5%; Pred. No. 2.70e+01;
 Matches 17; Conservative 9; Mismatches 15; Indels 1; Gaps 1;

Db 122 VKDLKIFTLXGLCPSEYELPDKFEEVPRIFKDLGYPPA 163
 QY 129 VK-FLKEIPGALVLVASYDDPGTKMNDSRKLFSLDGSYA 169

RESULT 9

ID W80950 standard; Protein; 642 AA.
 AC W80950;
 DT 16-FEB-1999 (first entry)
 DE Amino acid sequence of the human nuclear protein.
 KW Human nuclear protein; HEC; cell growth abnormality; hybridisation;
 KW amplification; antibody; immunoassay; malignancy; mitosis; antigen;
 KW immunoprecipitation; immunisation; vaccine.
 OS Homo sapiens.
 PN W09845433-Al.
 PD 15-OCT-1998.
 PF 03-APR-1998; U06727.
 PR 04-APR-1997; US-042609.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Allred DC, Channess GC, Clark GM, Hilsenbeck SG,
 PI Osborne CK;
 DR WPI; 98-594481/50.
 DR N-PSDB; V68588.
 PT New nucleic acid encoding human highly expressed in cancer nuclear
 PT protein - used for diagnosis and for modulation of the cell cycle to
 PT control malignant and other cell growth abnormalities
 PS Claim 2; Pages 58-61; 75pp; English.
 CC This is the amino acid sequence of a human nuclear protein (HEC) used
 CC in the method of the invention to diagnose cell growth abnormalities.
 CC fragments of HEC are used to detect HEC-encoding nucleic acid, in
 CC usual hybridisation or amplification assays, while antibodies are used
 CC in conventional immunoassays to detect HEC or peptides. HEC is used to
 CC modulate cell cycle progression (by disrupting chromatid separation)
 CC and antibodies are used to disorder sister chromatid alignment and
 CC separation in interphase cells, disrupting mitosis, particularly for
 CC control of malignancy and other cell growth abnormalities at the
 CC mitosis stage. Antibodies can also be used for protein purification, to
 CC isolate sequences encoding HEC or related proteins; to study HEC
 CC distribution in cells and to isolate antigens by immunoprecipitation.
 CC HEC may also be used in vaccines and antibodies for passive
 CC immunisation.
 SQ Sequence 642 AA;

Query Match 5.6%; Score 94; DB 1; Length 642;
 Best Local Similarity 40.5%; Pred. No. 2.70e+01;
 Matches 17; Conservative 9; Mismatches 15; Indels 1; Gaps 1;

Db 122 VKDLKIFTLXGLCPSEYELPDKFEEVPRIFKDLGYPPA 163
 QY 129 VK-FLKEIPGALVLVASYDDPGTKMNDSRKLFSLDGSYA 169

RESULT 10

ID R42381 standard; Protein; 146 AA.
 AC R42381;
 DT 19-APR-1994 (first entry)
 DE Haemin-binding protein produced as fusion protein in pGCH5.
 KW Haemophilus somnus; immunogenic; haemolysin; LppB; LppC;
 KW thromboembolic meningococcal septicemia; arthritis;
 KW pneumonia; lktA gene; Pasteurella haemolytica.
 OS Haemophilus somnus.
 PN W09321323-A.
 PD 28-OCT-1993.
 PF 05-APR-1993; CA0135.
 PR 09-APR-1992; US-865050.
 PR 04-JUN-1992; US-893424.
 PR 04-JUN-1992; US-893426.
 PR 29-MAR-1993; US-038287.
 PR 29-MAR-1993; US-038288.
 PR 29-MAR-1993; US-038719.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
 PI Rioux C, Theisen M;
 DR WPI; 93-351733/44.
 DR N-PSDB; Q51082.
 PT Haemophilus somnus immunogenic proteins used in vaccines -
 PT selected from haemin-binding protein, haemolysin, LppB and LppC,
 PT and corresp. DNA
 PS Claim 1; Fig 6; 119pp; English.

KW	thromboembolic meningoenkephalitis; septicaemia; arthritis;
OS	pneumonia; haemoglobin-binding protein.
KW	Haemophilus somnus.
OS	WO9321323-A.
PN	PD 28-OCT-1993.
PD	PD 05-APR-1993. CA0135.
PF	PD 09-APR-1992; US-865050.
PR	PR 04-JUN-1992; US-893424.
PR	PR 04-JUN-1992; US-893426.
PR	PR 29-MAR-1993; US-038287.

29-MAR-1993; US-038719.
PR (UYSA-) UNIV SASKATCHEWAN.
PA Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
PI

DR WPI; 93-351733/44.
DR N-PSDB; Q51080.
PT Haemophilus somnus immunogenic proteins used in vaccines -

CC A genomic cosmid library of *Haemophilus somnus* HS25 DNA was screened

activity. Positive clones were subcloned various times, resulting in plasmid pRP501, which binds haematin but is not haemolytic. The clone was sequenced and was found to contain several open reading frames, potentially encoding 8 proteins. The haematin binding protein (encoded by the hmb gene) was encoded by ORF1. The protein can be used in vaccines for preventing or treating *H. somnus* infections, which cause thromboembolic meningo-encephalitis, septicæmia,

[illegible]

RESULT	13	
ID	RI0957	standard; Protein; 191 AA.
AC	RI0957	
DT	16-APR-1991	(first entry)
DE	Bovine Somatotropin analogue #20.	
DE	bovine somatotropin; bST; dairy cow;	helical stability.
KS	Bos taurus.	
FW	key	Location/Qualifiers
FT	misc_difference	126
FT		/label= Ala, Cys, Leu, Gln, His, Lys, Val, Ile,
FT		Phe, Tyr, Trp, Thr, Gly, Ser, Asp, Asn,
FT		Pro, Arg
PN	WG9100870-A.	
PD	24-JAN-1991.	

27. UCR 1988; 005550.
PR 10-JUL-1989; US-377926.
PA (UPJO) UPJOHN CO.
PI Lehrman SR, Havel HA, Plaisted SM, Brems DN;

DK WPI; 91-031315/07.
PT Animal somatotropin analogues - have aminoacid changes at
PT residues 96 to 133 to reduce hydrophobicity or helical stability
PT of somatotropin

hydrophobicity and helical stability in the region of the mutation. The analogue-bst has a lower propensity of the partially denatured protein for aggregation and precipitation during processing. It also has enhanced bioactivity and can be used for enhancing the growth of an animal, e.g. beef cattle or increasing milk production in dairy cows. The mutation was introduced by standard site-directed mutagenesis.

QY	4	ccggcgaggttgcatatagaggaggcgctctgtga-gcactacccctccagcaactgg	62
Db	67	CTGGGCCACTGTCATCTAGAGGAGGCCGTCTGTGAGGCACCTACCCCTCCAGCAACTGG	126
QY	63	gaggtgggaactgtcaagaactgcccaggggtgtgtgtcagctgggttcaggggacctaacgc	122
Db	127	GAGTGGGAGCTGTCAAGCTGCCAGGGTGTGTCTAGCTGGGTTCAGGGGACCTACCGC	186
QY	123	acctgtggaacaacctgccttccatgaaacagggaagtgtggagcctcaagccctcg	182
Db	187	ACCTGTGGACCACTTCGCTTCCTCATGACAGCGGAAGTGGGAGCCTGAGCCCTCG	246

Db 421 CTTCTCTGACTTGGGAGTTCCTACGAAACAACTGGGCTCCGGGACAGCTGGGTCTT 480
 QY 761 catagagcaaaagactcaggggttaaaagccctttgacagcttcttaaaagacagccc 820
 Db 481 CATAGGAGCAAAAGACTCAGGGGTAAAGCCCTTTGAGCAGTCTTTAAAGAACAGCCC 540
 QY 821 agacacaaacatacagagggatggccagagctgctggagatggagggtgcgatcccc 880
 Db 541 AGACACAACAATACAGGGATGGCCAGAGCTGCTGGAGATGGAGGGCTGCATGCCCCC 600
 QY 881 gaagccattttaggggtggcttggctctctctcctcagccaggggctgaagaagctcctgcc 940
 Db 601 GAAGCCATTTTGGGTGGCTTGGCTCTTCTCAGCCAGGGGCTGAAGAAGCTCCTGCC 660
 QY 941 tgaacttagagtcagagccggcaggggtgagagagagagagaggggtgctgctgga 1000
 Db 661 TGACTTAGGAGTCAGAGCCCGGAGGGGCTGAGGAGGAGGAGGAGGGGTGCTGCTGGA 720
 QY 1001 aggtgctgaggttccttgacgctgtgtgcgctctctcctcctcctcgaacagaaacctcc 1060
 Db 721 AGGTGCTGAGGTCCTTGACGCTGTGTGCGGCTCTCTCCTCGGAAACAGAACCTCC 780
 QY 1061 cacagcaatctaccgccggaagaccagctcagaggggtcctctcttgaaacagctgctgt 1120
 Db 781 CACAGCACATCTTACCGGAAGACCACTCAGAGGCTCCTTGTGGAACAGCTGTCTGT 840
 QY 1121 ggagagaatgggtgtcttcgtaggagctgctgacgctgctgctgaggaagagcaaac 1180
 Db 841 GGAGAGAATGGGTGCTTGTGCTCAGGAGCTGCTGAGCGGTGCTGAGGAAGGAGCAAAAC 900
 QY 1181 tgccagacttgagcccaataataatttttttggctgga 1221
 Db 901 TGCCAGACTTGAGCCCAATTAATTTTATTTTGTGGTA 941

RESULT 3

V16671
 ID V16671 standard; cdna; 548 BP.
 AC V16671;
 DT 22-JUN-1998 (first entry)
 DE Polynucleotide sequence of a colon-specific gene.
 KW Colon-specific gene; probe; detection; expression; human;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..406
 FT /*tag= a
 PN US5733748-A.
 PD 31-MAR-1998.
 PF 06-JUN-1995; 469667.
 PR 06-JUN-1995; US-469667.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen C, Yu G;
 DR WPI; 98-229823/20.
 DR P-PSDB; W46878.
 PT Colon-specific nucleic acids - useful as probes for detecting colon cancer micrometastases
 PS Claim 15; Fig 4; 51pp; English.
 CC V16668-81 represent polynucleotide sequences of partial or full-length cDNA clones of colon-specific genes. The polynucleotides can be used as probes to detect expression of the corresponding human genes, e.g. in diagnostic assays for detecting micrometastases of colon cancer.
 CC Recombinant cells containing the polynucleotides can be used to produce the polypeptides, in order that antibodies can be raised and used in further screening or diagnostics.
 SQ Sequence 548 BP; 137 A; 128 C; 173 G; 107 T;

Query Match 42.2%; Score 515.6; DB 1; Length 548;
 Best Local Similarity 98.2%; Pred. No. 1e-141;
 Matches 540; Conservative 1; Mismatches 2; Gaps 2;
 QY 486 atgagctctgtgaaaaacaatgtggcagagggccttaaacatcgcccttggtgaatggaacc 545

Db 1 ATGAGTCTCTGTGAAAAACAATGTGGGAGAGGCTTAACATCGCCCTGGTGAATGGAACC 60
 QY 546 acggagagctgtgtggacagaagcatttgacatgtactctgagatgttatgcacctta 605
 Db 61 ACGGAGAGTGTGTGGGACAGAAGGCATTTGACATGTACTCTGGAGATGTTATGCACCTA 120
 QY 606 gtgaataatcccttaagaaattccgggggtgacactggtgctggtggctcctcactacagat 665
 Db 121 GTGAATTTCTTAAAGAAATTCGGGGGTGCATGGTGTGGTGGCTCCTACGACGAT 180
 QY 666 ccaggagacccaaatgaacatgaagcaggaaactctctctgacttggggagttcctac 725
 Db 181 CCAGGGACCAAAATGACGATGAAAGCAGGAAACTCTTCTCTGACTTGGGAGTTCCTAC 240
 QY 726 gcaaaacaactgggtctcgggacagctgggtcttcctataggagccaaagacctcaggggt 785
 Db 241 GCAAAACAACATGGGGCTTCCGGGACAGCTGGGTCTTCATAGGAGCCAAAGACCTCAGGGGT 300
 QY 786 aaagccctcttgagcagttcttaagaaacagccacagacacaaataacagagggatgg 845
 Db 301 AAAGCCCTTTGAGCAGTTCTTAAAGAACAGCCAGACACAAACAATACGAGGATGG 360
 QY 846 ccagagctgtgtgagatggagggtgcattgcccccgagccatttttaggggtggtgtggtg 905
 Db 361 CCAGAGCTGCTGGAGATGGAGGGCTGCATGCCGCCCGAAGCCATTTAGGGTGGCTGGC 420
 QY 906 tcttctcagccaggggctgaaagactcctgctgactgactgactgactgactgactgactg 965
 Db 421 TCTTCTCAGCCAGGGGCTTGAAGAAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 QY 966 gggctgagagagagagcaggggtgctgctgctgctgctgctgctgctgctgctgctg 1025
 Db 480 GGCTGNAGGAGGAGGAGGAGGGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538
 QY 1026 tgcgcgcct 1035
 Db 539 TGTGCGGCT 548

RESULT 4

T45883
 ID T45883 standard; cdna; 548 BP.
 AC T45883;
 DT 13-MAR-1997 (first entry)
 DE Human colon specific gene CSG4 cDNA partial clone.
 KW Colon specific gene; CSG4; colon cancer; metastasis; diagnosis;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..408
 FT /*tag= a
 PN WO9639419-A1.
 PD 12-DEC-1996.
 PF 06-JUN-1995; U07289.
 PR 06-JUN-1995; WO-U07289.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Yu G;
 DR WPI; 97-043054/04.
 DR P-PSDB; W06547.
 PT Human colon specific genes and their expression products - detection of which, in non-colon tissue samples, can be used as indication of colon cancer metastasis
 PS Claim 1; Fig 4; 60pp; English.
 CC 13 cDNA clones (T45880-92), most of them partial clones, correspond to human colon specific genes, designated CSG1, CSG2, etc., that are primarily expressed in tissues derived from the colon. CSG7 and CSG10 show reduced expression in colon cancer cells as compared to that in normal cells; the remaining genes are overexpressed in colon cancer. The partial cDNA sequences can be used to isolate full-length clones and genomic clones including the complete gene. CSG nucleic acids can be used to produce CSG polypeptides (see also WO6545-53) in transformed host cells, as probes to detect disorders


```

Oy 809 aaagaacagccagacacaaatacagggatggccagagctgctggagatggagg 868
    |||||||
Db 181 AAAGAAGCCAGACACAAACAAATACGAGGATGCCAGAGCTGCTGGAGATGGAGG 240
    |||||||
Oy 869 ctgcagatcccc 879
    |||||||
Db 241 CTGCATGCCCC 251

RESULT 9
V63510
ID V63510 standard; cDNA; 228 BP.
AC V63510;
DE 28-JAN-1999 (first entry)
DE Clone 1807758 EST corresponding to CS141 sequence.
KW EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
KW ulcerative colitis; pancreatitis; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9844133-A1.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABBO ) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI: 98-568280/48.
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
PT develop products for the diagnosis and treatment of e.g. cancers,
PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
PT pancreatitis.
PS Claim 1: Page 91; 116pp; English.
PS V63504-15 represent a set of contiguous and partially overlapping EST
PS sequences designated CS141. The sequences are isolated from a cDNA
PS library made from gastrointestinal tract tumour and normal tissues.
CC The CS141 gene is useful as a marker for gastrointestinal tract
CC disorders. The methods and products can be used in detecting, or
CC diagnosing, staging, monitoring, prognosticating, preventing, or
CC determining the predisposition to diseases and conditions of the
CC gastrointestinal tract, such as gastrointestinal tract cancers,
CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
CC Crohn's disease, ulcerative colitis, pancreatitis.
SQ Sequence 228 BP; 63 A; 48 C; 68 G; 49 T;

Query Match 18.7%; Score 228; DB 1; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.4e-57;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 495 gtgaaaaaataatggcagagccctaaacatcgccctggtgaatggaaccagggact 554
    |||||||
Db 1 GTGAAAACAATGTGGCGAGAGCCCTAAACATCGCCCTGCTGAATGGAACACCGGAGCT 60

Oy 555 gtgctgggacagaagcattgacatgtactctggagatgttatgcacctagtgaattc 614
    |||||||
Db 61 GTGCTGGGACAGAAGGCATTGTGACATGTACTCTGGAGATGTTATGCACCTAGTGAATTC 120

Oy 615 cttaaagaattccgggggtgacctggctggctggctctacagatccagggacc 674
    |||||||
Db 121 CTTAAAGAARTTCCGGGGGGTGCATGTGCTGGTGGGCTCTACGACGATCCAGGGACC 180

Oy 675 aaatgaacgatgaagcaggaactctctgacttggggagtcc 722
    |||||||
Db 181 AAAATGACGATGAAGACGAGGAATCTTCTCTGACTGGGGAGTTC 228

RESULT 10
V63507
ID V63507 standard; cDNA; 233 BP.
AC V63507;

```

```

DT 28-JAN-1999 (first entry)
DE Clone 775762 EST corresponding to CS141 sequence.
DE EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
KW ulcerative colitis; pancreatitis; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9844133-A1.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABBO ) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI: 98-568280/48.
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
PT develop products for the diagnosis and treatment of e.g. cancers,
PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
PT pancreatitis.
PS Claim 1: Page 90; 116pp; English.
PS V63504-15 represent a set of contiguous and partially overlapping EST
PS sequences designated CS141. The sequences are isolated from a cDNA
PS library made from gastrointestinal tract tumour and normal tissues.
CC The CS141 gene is useful as a marker for gastrointestinal tract
CC disorders. The methods and products can be used in detecting, or
CC diagnosing, staging, monitoring, prognosticating, preventing, or
CC determining the predisposition to diseases and conditions of the
CC gastrointestinal tract, such as gastrointestinal tract cancers,
CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
CC Crohn's disease, ulcerative colitis, pancreatitis.
SQ Sequence 233 BP; 44 A; 73 C; 77 G; 39 T;

Query Match 17.1%; Score 209; DB 1; Length 233;
Best Local Similarity 97.4%; Pred. No. 5.3e-52;
Matches 223; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Oy 4 ccggcagggctgcatctagagaggccgtctgtga-gccactaccctccagcaactgg 62
    |||||
Db 5 CTGGCCCACTGCATCTAGAGAGGGCGGCTGTGTAGGGCCACTACCCCTCCAGCACTGG 64

Oy 63 gagtgggactgtcagaagctggccagggtggtggtcagctgggtcagggacctacggc 122
    |||||||
Db 65 GAGGTGGGACTGTCAGAAGCTGGCCAGGCTGGTGGTGCAGCTGGGTGAGGACCTACGGC 124

Oy 123 acctgtggaccacctgcgcttctccatcgaagcagggaagtggagcctcgagccctcg 182
    |||||||
Db 125 ACCTGTGGACCACTCGCCTTCTCCATCGAAGCAGGGAAGTGGAGCCTCGAGCCCTCG 184

Oy 183 ggtggaagctgaccccaagccaccccttcacctgacagatgagagtgt 231
    |||||||
Db 185 GGTGGAGCTGACCCCAAGCCACCCTTCACCTGGACAGATGAGAGTGT 233

RESULT 11
V63506
ID V63506 standard; cDNA; 250 BP.
AC V63506;
DE 28-JAN-1999 (first entry)
DE Clone 2645837 EST corresponding to CS141 sequence.
KW EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
KW ulcerative colitis; pancreatitis; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9844133-A1.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABBO ) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,

```


CC gastrointestinal tract, such as gastrointestinal tract cancers,
 CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
 CC Crohn's disease, ulcerative colitis, pancreatitis.
 SQ Sequence 187 BP; 35 A; 42 C; 74 G; 33 T;

Query Match 14.7%; Score 179.2; DB 1; Length 187;
 Best Local Similarity 96.8%; Pred. No. 2.6e-43;
 Matches 181; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 839 gggatggccagagctgctggagatggagggtgcctgcatcccccgaagccatttaggggtg 898
 |||||
 DB 1 GGGATGGCCAGAGCTGCTGGAGATGGAGGCTGCATGCCCCGAGCCATTTTAGGTGG 60

QY 899 ctgtggcttctctctccagccaggggctggaagaagctcctgctgaacttaggaagcagagc 958
 |||||
 DB 61 CTGTGGCTCTTCTCAGCCAGGGGCTCGAAGAAGCTCCTGCCTGACTTAGGAGTCAAGC 120.

QY 959 ccggcaggggtgaggaagagacagaggggtgctgctggaaggtgctgacaggtccttg 1018
 |||||
 DB 121 CCGGcaggggtgaggaagagagcaggggtgctgctggaaggtgctgcaagtccttg 180

QY 1019 cagcgtc 1025
 |||||

DB 181 AAGNNG 187

RESULT 13
 V63505
 ID V63505 standard; cDNA; 250 BP.
 AC V63505;
 DT 28-JAN-1999 (first entry)
 DE Clone 1338704 EST corresponding to CS141 sequence.
 KW EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;
 KW gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
 KW ulcerative colitis; pancreatitis; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN W09844133-Al.
 PD 08-OCT-1998.
 PF 31-MAR-1998; U06337.
 PR 31-MAR-1997; US-828489.
 PA (ABEO) ABBOTT LAB.
 PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
 PI Gordon J, Granados EN, Hayden M, Hodges JC, Klass MR,
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 DR WPI: 98-568280/48.
 PT New gastrointestinal tract specific polynucleotides, CS141 - used to
 PT develop products for the diagnosis and treatment of e.g. cancers,
 PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
 PT pancreatitis.
 PS Claim 1; Page 89; 116pp; English.
 CC V63504-15 represent a set of contiguous and partially overlapping EST
 CC sequences designated CS141. The sequences are isolated from a cDNA
 CC library made from gastrointestinal tract tumour and normal tissues.
 CC The CS141 gene is useful as a marker for gastrointestinal tract
 CC disorders. The methods and products can be used in detecting,
 CC diagnosing, staging, monitoring, prognosticating, preventing, or
 CC determining the predisposition to diseases and conditions of the
 CC gastrointestinal tract, such as gastrointestinal tract cancers,
 CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
 CC Crohn's disease, ulcerative colitis, pancreatitis.
 SQ Sequence 250 BP; 44 A; 80 C; 84 G; 38 T;

Query Match 13.6%; Score 165.6; DB 1; Length 250;
 Best Local Similarity 95.2%; Pred. No. 2.9e-39;
 Matches 179; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 4 ccggcaggttgcatagagagggcgctgtgta-gccactaccctccagcaactgg 62
 |||||
 DB 63 CTGGCCNCACTGCACTCTAGAGGAGGCCGCTCTGTWAGGCCACTACCCCTCCAGCACTGG 122

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 17:30:02 ; Search time 57.43 Seconds
(without alignments)
2763.569 Million cell updates/sec

Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgccggcagggtgcatct.....aaattttattttgctggta 1221

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515.6	42.2	548	2	US-08-469-667-6
2	515.6	42.2	548	6	PCT-US95-07289-6
3	46.6	3.8	7218	1	US-08-232-463-14
4	38	3.1	6803	5	US-08-665-259-19
5	35.8	2.9	3475	2	US-07-960-389-1
6	34.4	2.8	2363	2	US-07-923-724-7
7	34.4	2.8	2363	3	US-08-609-426A-7
8	34.4	2.8	2379	3	US-08-374-652C-1
9	33.8	2.8	3240	2	US-08-294-189-2
10	33.6	2.8	2085	3	US-08-283-917-8
11	33.6	2.8	2085	3	US-08-961-716-8
12	33	2.7	1404	1	US-08-151-574-33
13	33	2.7	1404	1	US-08-146-424-19
14	33	2.7	1404	3	US-08-693-709-1
15	33	2.7	1404	3	US-08-419-448-33
16	33	2.7	6756	1	US-08-151-574-31
17	33	2.7	6756	3	US-08-419-448-31
18	33	2.7	31571	1	US-08-323-443B-1
19	32.6	2.7	579	2	US-08-190-560-3
20	32.6	2.7	579	2	US-08-469-277-3
21	32.6	2.7	579	3	US-08-468-946-3
22	32.6	2.7	579	4	US-08-468-942-3
23	32	2.6	2923	2	US-08-377-292-6
24	32	2.6	2923	3	US-07-989-847-7
25	32	2.6	2923	7	5187076-5
26	32	2.6	5057	3	US-08-365-486A-12
27	32	2.6	5108	1	US-07-642-002-1

28	31.8	2.6	340	5	US-08-888-077A-31	Sequence 31, Appl
29	31.6	2.6	1947	1	US-07-951-715A-3	Sequence 3, Appl
30	31.6	2.6	1947	3	US-08-459-448A-3	Sequence 3, Appl
31	31.6	2.6	1947	5	US-08-459-595A-3	Sequence 3, Appl
32	31.6	2.6	3468	1	US-07-951-715A-4	Sequence 4, Appl
33	31.6	2.6	3468	1	US-07-951-715A-8	Sequence 8, Appl
34	31.6	2.6	3468	3	US-08-459-448A-4	Sequence 4, Appl
35	31.6	2.6	3468	3	US-08-459-448A-8	Sequence 8, Appl
36	31.6	2.6	3468	5	US-08-459-595A-4	Sequence 4, Appl
37	31.6	2.6	3468	5	US-08-459-595A-8	Sequence 8, Appl
38	31.6	2.6	3546	1	US-07-951-715A-10	Sequence 10, Appl
39	31.6	2.6	3546	1	US-07-951-715A-12	Sequence 12, Appl
40	31.6	2.6	3546	1	US-07-951-715A-14	Sequence 14, Appl
41	31.6	2.6	3546	1	US-07-951-715A-27	Sequence 27, Appl
42	31.6	2.6	3546	3	US-08-459-448A-10	Sequence 10, Appl
43	31.6	2.6	3546	3	US-08-459-448A-12	Sequence 12, Appl
44	31.6	2.6	3546	3	US-08-459-448A-14	Sequence 14, Appl
45	31.6	2.6	3546	3	US-08-459-448A-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-469-667-6
; Sequence 6, Application US/08469667
; Patent No. 573748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrnie, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..405
US-08-469-667-6

Query Match 42.2%; Score 515.6; DB 2; Length 548;


```
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
;
US-08-232-463-14
```

```
Query Match 3.8%; Score 46.6; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred. No. 0.0006;
Matches 25; Conservative 213; Mismatches 177; Indels 0; Gaps 0;

Qy 480 atgatacagtcctgtgaaacaacatgtggcagagcctaaacatcgccctgggtgaat 539
| || | : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1447 AAGAATTGGTACRrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1388

Qy 540 ggaaccagggagctgtgctgggacagaagcattgacatgtactctggagatgtatg 599
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1387 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1328

Qy 600 cacctagtgaattccttaaagaattccggggggtgactgtgctgtggtgctctctac 659
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1327 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1268

Qy 660 gacatccaggggacaaatgaacagtgaagcaggaaactcttctgacttggggagt 719
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1267 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1208

Qy 720 tcttacgcaaaactgggtctccgggacagctgggtcttctatagagcaaacactc 779
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1207 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1148

Qy 780 aggggtaaagcccttttgacagtctctaaagaacagccagacacacaaatacag 839
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1088

Qy 840 ggatggccagactgctggagatggaggtgcatgcccccgaaagccatttttag 894
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1087 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1033
```

```
RESULT 4
US-08-665-259-19
; Sequence 19, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-665-259-19

Query Match 3.1%; Score 38; DB 5; Length 6803;
Best Local Similarity 59.4%; Pred. No. 0.19; Mismatches 1; Gaps 1;
Matches 82; Conservative 0; Indels 55;

Qy 868 gtgcatgcccccgaaagccattttagggtggctgtggtctctctcagccagggcctga 927
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 GCTCCCTGCGGCTAGACTGCCAGCCCATCTATGCCCTCTCCCAGCCTGTGCCCAA 217

Qy 928 agaagctctgctgacttagagtcagagccggcaggggctgagggagagagcaggg 987
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 AGCTGAGAGTCCACATC-TAGGGGTGAGGGGTGGGGAGGGAGGCGAAGCACATG 276

Qy 988 ggtgctgctggaaggtg 1005
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 CGCCTGAGTTGCAGGTG 294

RESULT 5
US-07-960-389-1
; Sequence 1, Application US/07960389
; Patent No. 5705611
; GENERAL INFORMATION:
; APPLICANT: HAYASHIDA, Kasuhiro;
; TITLE OF INVENTION: Human GM-CSF Receptor Component
; NUMBER OF SEQUENCES: 2
```

```

Query Match          2.94; Score 35.8; DB 2; Length 3475;
Best Local Similarity 61.1%; Pred. No. 0.61;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 791 ccccttgagcagtcctcaagaacagccagacacacaaatacagggatggcgaga 850
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1016 CTCCTTTGGCCTATTCTACAAAGCCAGCCAGCCAGATGTCAGGGAGAGAGTGTCCCCAGT 1075
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 851 gctgctggatggagggctgcagtcgcccccgaaagc 885
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1076 GCTGAGGGAGGGGCTCGGCAGCCTCCACACCAAGGC 1110
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-07-923-724-7
: Sequence 7, Application US/07923724
: Patent No 5780292
: GENERAL INFORMATION:
: APPLICANT: Nevalainen, Helena K.M.
: APPLICANT: Paloheimo, Marja T.
: APPLICANT: Miettinen-Oinonen, Arja S.K.
: APPLICANT: Torkkeli, Tuula K.
: APPLICANT: Cantrell, Michael
: APPLICANT: Piddington, Christopher S.
: APPLICANT: Rambossek, John A.
: APPLICANT: Turunen, Marja K.
: APPLICANT: Fagerström, Richard B.
: TITLE OF INVENTION: Production of Phytase Degrading Enzymes
: TITLE OF INVENTION: in Trichoderma
: NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox

```

```

; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/923,724
; FILING DATE: 31-JUL-1992
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 8610600
; FILING DATE: 30-APR-1986
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1050.0240004
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2363 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: join(404..447, 550..1906)
;
US-07-923-724-7
;
;
; Query Match 2.8%; Score 34.4; DB 2; Length 2363;
; Best Local Similarity 48.0%; Pred. No. 1.3;
; Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps
;
QY 252 gccctcatcttggccatgacacatggtgctgcctgcgcacaaaggagatccaggttaaa 371
Db 989 GGCTCCAGCGGGGTGATCGCCTCCGCGAGAAATTCATTGAGGGCTTCACAGACCAAG 1048
QY 312 atgaaacaccatccgtctgcacgtggctggcctgcgcacaaaggagatccaggttaaa 371
Db 1049 CTGAAGGATCTCGTGCCCGACGGGGCCAATGTCGCCCAAGATCGACGTGGTCATTTCC 1108
QY 372 aagtacaagtgtggcctcatcaagccctgccacgccaactacttgcgttttaaaactgc 431
Db 1109 GAGGCCAGTCTATCCNACACACTCTCGNACCAGCCACTGCTGTTGAACAGACG 1168
QY 432 agtggggccgcacaaagtctggtggc 455
Db 1169 GAATTGGCCGATACGTCGAAGCC 1192
;
;
RESULT 7
US-08-609-426A-7
; Sequence 7, Application US/08609426A
; Patent No. 5830733
;
; GENERAL INFORMATION:
;
; APPLICANT: Nevalainen, Helena K.M.
; APPLICANT: Paloheimo, Marja T.
; APPLICANT: Miettinen-Oinonen, Arja S.K.
; APPLICANT: Torkkeli, Tuula K.
; APPLICANT: Cantrell, Michael

```

```

; APPLICANT: Piddington, Christopher S.
; APPLICANT: Rambosek, John A.
; APPLICANT: Turunen, Maria K.
; APPLICANT: Fagerstr m, Richard B.
; APPLICANT: Houston, Christine S.
; TITLE OF INVENTION: Production of Phytase Degrading Enzymes
; TITLE OF INVENTION: in Trichoderma
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,426A
; FILING DATE: 01-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/923,724
; FILING DATE: 31-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 8610600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Grant E.
; REGISTRATION NUMBER: P-41,264
; REFERENCE/DOCKET NUMBER: 1050.0080001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2363 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(404..447, 550..1906)
; US-08-609-426A-7

```

```

Query Match      2.8%; Score 34.4; DB 3; Length 2363;
Best Local Similarity 48.0%; Pred. No. 1.3;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 252 gcctcatcttggccatagtcacgacatggtttatttcgaagtcacatgagcttcagc 311
Db 989 GGTCTCAGCGCGGTGATCGCTCCGGGAGAAATTCATTGAGGCGCTCCAGAGCACCACAG 1048

QY 312 atgaataaccatcgctgtgcacgctggctgcctcccccacacgaagagatccagggttaa 371
Db 1049 CTGAAGGATCTCGTGCCCGCCGCGGCAATCGTCCCAAGATCGACGTGGTCAATTCC 1108

QY 372 aagtacaagtgtggcctcatcaagccctgccagcgaactacttttcggtttaaatctgc 431
Db 1109 GAGGCCAGCTCATCCACAACACTCTCGACCCAGCACCTGCTACTGTCTTTGAAGACAGC 1168

QY 432 agtggggccgccaacgtcgtgggc 455
      ||||| | ||||| |

```

```

Db 1169 GAATTGGCGGATACCGTCGAAGCC 1192

RESULT      8
US-08-374-652C-1
; Sequence 1, Application US/08374652C
; Patent No. 5834286
; GENERAL INFORMATION:
; APPLICANT: NEVALAINEN, HELENA K.M.
; APPLICANT: PALOHEIMO, MARJA T.
; APPLICANT: FAGERSTROM, RICHARD B.
; APPLICANT: MIETTINEN-OINONEN, ARJA S.
; APPLICANT: TURUNEN, MARJA K.
; APPLICANT: RAMBOSEK, JOHN A.
; APPLICANT: PIDDINGTON, CHRISTOPHER S.
; APPLICANT: HOUSTON, CHRISTINE S.
; APPLICANT: CANTRELL, MICHAEL A.
; TITLE OF INVENTION: RECOMBINANT CELLS, DNA CONSTRUCTS,
; TITLE OF INVENTION: VECTORS AND METHODS FOR EXPRESSING PHYTASE DEGRADING
; TITLE OF INVENTION: ENZYMES IN DESIRED RATIOS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,652C
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07058
; FILING DATE: 27-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925,401
; FILING DATE: 31-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1050.071001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(404..447, 550..1906)
; US-08-374-652C-1

```

```

Query Match      2.8%; Score 34.4; DB 3; Length 2379;
Best Local Similarity 48.0%; Pred. No. 1.3;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 252 gcctcatcttggccatagtcacgacatggtttatttcgaagtcacatgagcttcagc 311
Db 1005 GCTCCAGCGCGGTGATCGCTCCGGGAGAAATTCATTGAGGCGCTCCAGAGCACCACAG 1064

```


Qy	122	caactgtgtgaccacctcgcttcttcacatcgaagcagggaagtggagcctc	173
Db	52	GGCGTGTCCCGCCCTCTCCCGTGTCTCTCACTCAACGGCCGTCGCGGGCGC	1
RESULT	11		
US-08-961-716-8/C			
Sequence 8, Application US/08961716			
Patent No. 5880272			
GENERAL INFORMATION:			
APPLICANT: ADACHI, HIDEKI			
APPLICANT: TSUJIMOTO, MASAFUMI			
APPLICANT: INOUE, KEIZO			
APPLICANT: ARAI, HIROYUKI			
TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME			
TITLE OF INVENTION: AND GENE THEREOF			
NUMBER OF SEQUENCES: 31			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &			
ADDRESSEE: NEUSTADT, P.C.			
STREET: 1755 S. Jefferson Davis Highway, Suite 400			
CITY: Arlington			
STATE: Virginia			
COUNTRY: U.S.A.			
ZIP: 22202			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/961,716			
FILING DATE:			
CLASSIFICATION:			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/283,917			
FILING DATE: 03-AUG-1994			
APPLICATION NUMBER: JP 209943/1993			
FILING DATE: 03-AUG-1993			
ATTORNEY/AGENT INFORMATION:			
NAME: Oblon, No. 5880272man F.			
REGISTRATION NUMBER: 24,618			
REFERENCE/DOCKET NUMBER: 2292-030-0			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (703) 413-3000			
TELEFAX: (703) 413-2220			
TELEX: 248855 OPAT UR			
INFORMATION FOR SEQ ID NO: 8:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 2085 base pairs			
TYPE: nucleic acid			
STRANDEDNESS: unknown			
TOPOLOGY: unknown			
MOLECULE TYPE: cDNA			
ORIGINAL SOURCE:			
ORGANISM: Bos taurus			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 844..2073			
US-08-961-716-8			
Query Match	2.8%;	Score 33.6;	DB 3; Length 2085;
Best Local Similarity	56.2%;	Pred. No. 2.1;	
Matches	63; Conservative	0; Mismatches	49; Indels 0; Gaps 0;
Qy	62	ggaggtgggactgtcagaagtcgcccagggtgtgtgtcagctgggtcagggaacctacgg	121
Db	112	GGCCCTGTGCACCCCGACAAACCCGGCCAGGAGGAGGGGAGCGAGCGCGCCCTACCC	53
Qy	122	caactgtgtgaccacctcgcttcttcacatcgaagcagggaagtggagcctc	173

```

Qy  432 agtggggccgcacacgtcgtggccctactatgtgtttgaagacgcgatcatgagt 491
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  664 GAATTCGGCGATACCGTCGAAGCAATTTACCGCCACGTTCTGCTCCCTCCATTCGTCAA 723
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  492 cctgtgaaaaacaatgtgggcagag 516
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  724 CGTCTGGAGAACGACCTGTCGGTG 748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-146-424-19
; Sequence 19, Application US/08146424
; Patent No. 5593963
; GENERAL INFORMATION:
; APPLICANT: VAN OOIJEN, ALBERT J. J.
; APPLICANT: RIETVELD, KRIJN
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: PEN, JAN
; APPLICANT: SIJMONS, PETER C.
; APPLICANT: VERHOED, TEUNIS C.
; TITLE OF INVENTION: THE EXPRESSION OF PHYTASE IN PLANTS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/146,424
; FILING DATE: 02-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNEDY, BILL
; REGISTRATION NUMBER: 33,407
; REFERENCE/DOCKET NUMBER: 44615-20011.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1401
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 70
; US-08-146-424-19

```

	Query Match	2.78;	Score 33;	DB 1;	Length 1404;
	Best Local Similarity	45.3%;	Pred. No. 2.6;		
	Matches 120;	Conservative 0;	Mismatches 145;	Indels 0;	Gaps 0;
QY	252	gcctcattcttgcctatgttcacgacatgcatgtttatttcgaagctacatgagcttcagc	311		
DB	484	GGCTCCAGCCGGTGATCGCTCCGGCAAGAAATTCATCGAGGCTTCAGAGCACCAAG	543		
QY	312	atgaaaccatccgtctgtgccacgtgctgcgcctgcacacaaaggagatccaggtttaa	371		
DB	544	CTGAAGGATCTCTGTGCCACCGCGGCAATCTGCGCCAAAGATCGACGTGTCATTTC	603		

```

QY 372 agtacaagtgtgacctatcaagccctgcggccccgaacactactttcgctttaaaattcgc 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 604 GAGGCCAGCTCATCCAAACAACACTCTCGAACCCGAGCACCTGTCTTCGAAGACAGC 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 agtggggccgccacgctcggtgggcccctactatgtgtttgaaagaccgcgatcatcagtg 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 664 GAATGGCCGATACCGTCAAGCCAATTTCACGCCACGTTCTGTCGCCCTCCATTCGTCAA 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 492 cctgtgaaaaaacatgtgggcagag 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 724 CGTCTGGAGAAGACCTGTCCGGTG 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-693-709-1
; Sequence 1, Application US/08693709
; Patent No. 5770413
; GENERAL INFORMATION:
; APPLICANT: VAN OOLJEN, ALBERT J.J.
; APPLICANT: RIETVELD, KRIJN
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: PEN, JAN
; APPLICANT: SIJMONS, PETER C.
; APPLICANT: VERWOERD, TEUNIS C.
; TITLE OF INVENTION: THE EXPRESSION OF PHYTASE
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/693,709
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/146,424
; FILING DATE: 02-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24615-20011.10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1401
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: 70...1401
; OTHER INFORMATION:
; NAME/KEY: Signal sequence
; LOCATION: 1...72
; OTHER INFORMATION:
; IS-08-693-709-1

```

Query Match 2.7%; Score 33; DB 2; Length 1404;
Best Local Similarity 45.3%; Pred. No. 2.6;
Matches 120; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
QY 252 gccctcatcttgccatgctacgacatggtttatttcgaagctacatgagcttcagc 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 GGTCTCAGCGCGGTATCGCTCCCGCAAGAAATTCATCGAGGGGTTCCAGAGACCAAG 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 atgaaaccatccgtgcccacgctggctggcctgcccacacaaaggagatccaggttaaa 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 CTGAAGGATCCTCGTCCCGCAGCCGSCCAATCGTCCCAAGATCGACGTGTCATTTC 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 aagtaacagtgtggtcctatcaagccctgcccagcaactatttgcgtttaaaaatctgc 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 604 GAGGCCAGCTCATCCAAACACACTCTCGACCCAGGACCTGACCTGTCTTCGAAGACAGC 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 agtggggccgccaactgctggcctactatgtgtttgaagaccgcatgatcatgagt 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 GAATTGGCGATACCGTGAAGCCAAATTCACGCCACGTTCTGTCCTCCATTCGTCAA 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

US-08-419-448-33
Sequence 33, Application US/08419448
Patent No. 5863533

GENERAL INFORMATION:

APPLICANT: Robert F.M. Van Gortcom
APPLICANT: Willem Van Hartingsveldt
APPLICANT: Petrus A. Van Paridon
APPLICANT: Annemarie E. Veenstra
APPLICANT: Rudolf G.M. Luttin
APPLICANT: Gerardus Seltin
TITLE OF INVENTION: Cloning and Expression of Microbial
TITLE OF INVENTION: Phytase
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
SOFTWARE:

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/419,448
FILING DATE: 10-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 24615-20026.10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:

LENGTH: 1404 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Aspergillus ficuum (Aspergillus niger)
STRAIN: NRRL 3135
US-08-419-448-33

Query Match 2.7%; Score 33; DB 3; Length 1404;
Best Local Similarity 45.3%; Pred. No. 2.6;
Matches 120; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
QY 252 gccctcatctttgccaatgctcacgacatggtttatttcgaagctacatgagcttcagc 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 GGTCTCAGCGCGGTATCGCTCCCGCAAGAAATTCATCGAGGGGCTTCAGAGACCAAG 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 atgaaaccatccctctgcccacgctggctggcctgcccacacaaaggagatccaggttaaa 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 CTGAAGGATCCTCGTGGCCCGCAGCCGSCCAATCGTCCCAAGATCGACGTGTCATTTC 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 aagtacaagtgtggcctcatcaagccctgcccagcccaactactttgcgtttaaaaatctgc 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 604 GAGGCCAGCTCATCCAAACACACTCTCGACCCAGGACCTGACCTGTCTTCGAAGACAGC 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 agtggggccgccaactgctggcctactatgtgtttgaagaccgcatgatcatgagt 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 GAATTGGCGGATACCGTGAAGCCAAATTCACGCCACGTTCTGTCCTCCATTCGTCAA 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 492 cctgtgaaaaaacaatgtgggcagag 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 CGTCTGGAGAACGACCTGTCCGGTG 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 31, 2000, 19:14:15
Job time: 6253 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 16:54:03 ; Search time 799.5 Seconds
(without alignments)
6190.101 Million cell updates/sec

Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgcccgggcaggtgcactc.....aaattttattttgctggta 1221

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 9714632
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_est45:*
- 80: gb_est46:*
- 81: gb_est47:*
- 82: gb_gss1:*
- 83: gb_gss2:*
- 84: gb_gss3:*
- 85: gb_gss4:*
- 86: em_gss1:*
- 87: em_gss2:*
- 88: em_gss3:*
- 89: em_gss4:*
- 90: gb_gss5:*
- 91: gb_gss6:*
- 92: gb_gss7:*
- 93: gb_gss8:*
- 94: gb_gss9:*
- 95: em_gss5:*
- 96: em_gss6:*
- 97: em_gss7:*
- 98: em_gss8:*
- 99: em_gss9:*
- 100: em_gss10:*
- 101: em_gss11:*
- 102: gb_gss10:*
- 103: gb_gss11:*
- 104: em_gss12:*
- 105: gb_gss12:*
- 106: gb_gss13:*
- 107: gb_gss14:*
- 108: gb_gss15:*
- 109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
C 1	624.8	51.2	645	79	AW300770	AW300770 xk05q02.x
C 2	616	50.5	668	63	AW001287	AW001287 wf27e06.x
C 3	595.6	48.8	615	50	AW1691072	AW1691072 wf22c03.x
C 4	568.4	46.6	646	62	AW1936111	AW1936111 wo61h03.x
C 5	558	44.1	564	36	AA622758	AA622758 np76c04.s
C 6	522.2	42.8	538	44	AI245950	AI245950 qk45b04.x
C 7	518	42.4	518	35	AA563933	AA563933 nk23b02.s
C 8	498	40.8	498	45	AI348065	AI348065 qp56d09.x
C 9	496.8	40.7	594	35	AA552519	AA552519 nk14e01.s
C 10	495.4	40.6	508	60	AI801582	AI801582 tq91c11.x
C 11	495	40.5	518	36	AA622120	AA622120 ng56b03.s
C 12	471.6	38.6	479	63	AW001308	AW001308 wu28c10.x
C 13	464	38.0	464	39	AA847242	AA847242 of01b11.s
C 14	456.2	37.4	461	36	AA622570	AA622570 np21f04.s
C 15	449.4	36.8	454	35	AA552362	AA552362 nk04h09.s
C 16	449	36.8	455	49	AI660557	AI660557 we68b05.x
C 17	424	34.7	530	64	AW050790	AW050790 wz21e07.x
C 18	402	32.9	417	35	AA582787	AA582787 np35b01.s
C 19	401.8	32.9	443	63	AW000826	AW000826 wu45d09.x
C 20	371.4	30.4	396	36	AA643708	AA643708 nl96f08.s
C 21	370	30.3	375	51	AW1732367	AW1732367 nf61e11.x
C 22	356.4	29.2	434	39	AA857546	AA857546 of64c07.s
C 23	355	29.1	360	34	AA514424	AA514424 nf61e11.s
C 24	355	29.1	364	36	AA643616	AA643616 ng79b08.s
C 25	352	28.8	358	50	AI673534	AI673534 we75b10.x
C 26	320.8	26.3	325	35	AA543029	AA543029 nf92e04.s
C 27	310.4	25.4	321	61	AI821215	AI821215 nl612h10.y
C 28	307	25.1	337	27	AA025434	AA025434 ze84f10.s
C 29	302	24.7	306	31	AA297147	AA297147 EST112731
C 30	286	23.4	301	34	AA470683	AA470683 nel2h10.s
C 31	283.8	23.2	302	51	AI732198	AI732198 nel2h10.x
C 32	274.4	22.5	289	48	AI582013	AI582013 ar98b03.x
C 33	267	21.9	268	31	AA297176	AA297176 EST112740
C 34	256.2	21.0	417	31	AA298484	AA298484 EST114112
C 35	246	20.1	249	51	AI749731	AI749731 at32a08.x
C 36	230	18.8	461	36	AA611725	AA611725 v089f11.r
C 37	223	18.3	379	80	AA315283	AA315283 12472 MAR
C 38	216.2	17.7	377	27	AA025433	AA025433 ze84f10.r
C 39	209.4	17.1	387	37	AA451058	AA451058 vf86b07.r
C 40	181	14.8	778	62	AI883327	AI883327 fc60e01.y
C 41	180.6	14.8	525	80	AA323823	AA323823 u077a04.y
C 42	176.8	14.5	264	38	AA793766	AA793766 vf96e04.r
C 43	160.8	13.2	311	80	AA315346	AA315346 12597 MAR
C 44	151.6	12.4	410	24	NA1338	NA1338 yw68q05.r1
C 45	151.4	12.4	509	34	AA512284	AA512284 vj19g03.r

ALIGNMENTS

```

RESULT 1
LOCUS AW300770/c 645 bp mRNA EST 18-JAN-2000
DEFINITION xk05q02.x1 NCI CGAP_Col9 Homo sapiens CDNA clone IMAGE:2665874 3'
            similar to SM:EF07_MOUSE P97805 PROTEIN EF-7 ;, mRNA sequence.
ACCESSION AW300770.1 GI:6710447
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 645)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
        Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jul 8, 1999 this sequence version replaced gi:5420873.
        Contact: Robert Strausberg, Ph.D.
        Tel: (301) 496-1550
        Email: Robert_Strausberg@nih.gov

```

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/brp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 403.

FEATURES

source
Location/Qualifiers
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2665874"
/clone_lib="NCI_CGAP_Col9"
/tissue_type="moderately differentiated adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Normalized to Cot 50. Average insert size 1.32kb. Normalized version of NCI_CGAP_Col8. Library constructed by Life Technologies."
BASE COUNT 137 a 192 c 159 g 156 t 1 others
ORIGIN

Query Match 51.2%; Score 624.8; DB 79; Length 645;
Best Local Similarity 98.0%; Pred. No. 3.9e-160;
Matches 632; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 576 gacatgactctgagatgttatgcacactagtgaataatctctaaagaatactcgggggggt 635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 GACATGATTCTGAAGATGTATGCACATAGTGAATTTCTAAAGAAATTCGCGGGGT 586

QY 636 gcactggtgctgggtgctcctacgacgtatccaggagaccacaaatgaacgatgaacgag 695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 GCACGTGCTGTGGTGCTCTTACGCCGATCCAGGACCAAAATGACGATGAAGACAGA 536

QY 596 aaactctctgactctggtggagttctactgcgaacaaactcggggtccggggacagtgg 755
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 AAACCTCTCTGACTTTGGGGAGTTCCTACGCAAAAACAACTGGGCTTCCGGGACAGCTGG 466

QY 756 gtcttcataagaccacaaagacctcagggttaaaagccctttgagcaggtctctaaagaac 815
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 465 GTTTCATAGGAGCCAAAGACCTCAGGGGTAAAGCCCTTTGAGCAGTTCCTTAAAGAAC 406

QY 816 agccacagacacacaaataacgaggggagtgccagagctgctggagatgagggctgcatg 875
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 AGCCACAGACACAAACAAATACGAGGGATGCCAGAGCTGCTGGAGATGGAGGGCTGCATG 346

QY 876 cccccaagaccattttaggtggtgctgtgctctctcagccagggggcctgaagaagctc 935
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 CCCCCGAAGCCATTTTAGGTGGCTGTGGCTCTTCTCAGCCAGGGGCTGGAAGAAGCTC 286

QY 936 ctgctgacttagagtcagagcccgaggggctcagagagagagagagagaggggtctac 995
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 285 CTGCCTGACTTAGAGTCAGAGCCCGGCGGCTCAGGAGGAGGAGGAGGGGCTGCTGC 226

QY 996 gtggaaggtgctcaggtctctgcacgtgtgtcgcgcctctctcctcctcggaacagaac 1055
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 225 GTGGAAGGTGCTGCAGGTCTCTTGCACGCTGTGTCGCGCTCTCTCTCTCGAAACAGAAC 166

QY 1056 cctccacagacacactcctaccgccgaagacagcctcagaggggtcctcttggaaccagctg 1115
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 CCTCCACACACATCTTACCCGGAAGACAGCCTCAGAGGGTCTCTTCTGGAACACAGCTG 106

QY 1116 tctgtgagagaatgggtgcttcttcgtcagggactgctgacggtggtcctcaggaaga 1175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 TCTGTGGAGAGATGGGTGCTTTCGTCAGGGGACTCTGACGGCTGCTCTGAGGAGGA 46

```

QY 1176 caaactgcccagacttgagcccaataaatttttttttgcgtgt 1220
|||||
Db 45 CAAACTGCCAGACTTGAGCCCAATTAATTTATTTTGTGCTGT 1

RESULT 2
AW001287/c 668 bp mRNA EST 27-OCT-1999
LOCUS wu27e06.x1 Soares_Dieckgraefe_colon_NHUC Homo sapiens cDNA clone
DEFINITION IMAGE:2521282 3' similar to SW:EF07_MOUSE P97805 PROTEIN EF-7
;contains MSR1.t2 MBR22 repetitive element ;, mRNA sequence.
ACCESSION AW001287
VERSION AW001287.1 GI:5848203
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 668)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1996 this sequence version replaced gi:1132811.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 478.

FEATURES
source
1..668
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2521282"
/clone_lib="Soares_Dieckgraefe_colon_NHUC"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAGCTGGAGCGCGCCGCTTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@lm.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 142 a 197 c 170 g 159 t

ORIGIN
Query Match 50.5%; Score 616; DB 63; Length 668;
Best Local Similarity 96.8%; Pred. No. 1e-157;
Matches 639; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

QY 562 gacagaaggcatttgacattgactgtgagatgttatgacactagtgaattcttaag 621
|||||
Db 667 GCCAGCAGGCATTTGACATGCACTTAGGAGAGTATTGCACTTAGTCAATTAAGA 608
QY 622 aaattccgggggtgactgtgctgctccctacacgacatccaggaccacaatga 681
|||||
Db 607 TATTTCGCGGGGFGCACTGGTGTGTGGGCTCCCTACGACGATCCAGGCCACCAATGA 548
QY 682 acgatgaagcaggaactctctgacttggggagttctactcgaaaaaactgggct 741
|||||
Db 547 TCGATGAAGCAGGAAACTCTTCTCTGACTTGGGAGTTCTTACGAAACAACTGGGCT 488

QY 742 tcgaggacacactgggtcttccatagggagccaaagacactcaggggtaaaaagccctttgac 801
|||||
Db 487 TCCGGGACAGCTGGGTCTTCATAGGAGCCAAAGACCTCAGGGGTAAAGCCCCCTTTGAGC 428
QY 802 agttcttaagaagacagcccagacacacaaataacagaggatgcccagagctgctggaga 861
|||||
Db 427 AGTTCTTAAGAGACAGCCAGACACCAACAAATACGAGGGATGCCAGAGCTCCTGGAGA 368
QY 862 tggagggtcgtatgccccccgaagccatttttaggggtgctgtgcttctcttcagaccaggg 921
|||||
Db 367 TGGAGGGCTCATGCCGCCCGAAGCCATTTTAGGGTGGCTGTCTCTCTCAGCCAGGG 308
QY 922 gctggaagaagctcctcctcctgac-ttaggagtcagagcccgaggggctgagagaggg 980
|||||
Db 307 GCCTGAAGAAGCTCCTGCCCTTACTTTAGGAGTCAGAGCCCGGAGGGGCTGAGAGAGG 248
QY 981 agcaggggtgctgctgctggaaggtgctgcaggtcctctgcacgctgtgtcgcctcctcct 1040
|||||
Db 247 AGCAGGGGTGCTGCTGGTGGAGGTGCTGCAGGTCTTGCACGCTGTGTCGCGCTCTCCT 188
QY 1041 cctcgaaacagaaacctccacagcacatcctaccocgggaagaccacacctcagagggtcc 1100
|||||
Db 187 CCTCGAAACAGAAACCTCCTCCACAGCACATCTACCCGGAAGACCCAGCTCAGAGGGTCC 128
QY 1101 ttctggaaccagctgtctgtggagagaatgggtgcttctcaggaactgctgacggtc 1160
|||||
Db 127 TTTTGGAAACCAAGCTGTCTGTGGAGAGATGGGTGCTTTCTCAGGGACTGCTGACGGCT 68
QY 1161 ggtcctgagggaagacaaactgccagacttgagcccaataaatttttttgcgtgt 1220
|||||
Db 67 GGTCTGAGGAAGGACAAACTGCCAGACTTGAGCCCAATTAATTTATTTTGTGCTGCT 8

RESULT 3
AI691072/c

LOCUS AI691072 615 bp mRNA EST 18-DEC-1999
DEFINITION w122c03.x1 Soares_Dieckgraefe_colon_NHUC Homo sapiens cDNA clone
IMAGE:2351332 3' similar to SW:EF07_MOUSE P97805 PROTEIN EF-7 ;,
mRNA sequence.

ACCESSION AI691072
VERSION AI691072.1 GI:4902374
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 615)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1134301.
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 893 Std Error: 0.00
Seq primer: -40UP from Gibco

High quality sequence stop: 476.
Location/Qualifiers

1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2351332"

/clone_lib="Soares_Dieckgraefe_colon_NHUC"
/tissue_type="colonic mucosa from 5 ulcerative colitis
patients"
/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',

Db 285 TGCTGACTTAGGAGTAAAGCCGGCAGGCGCTCAGGAGGAGGAACAAGGGGTGCTGCG 226
Qy 997 tggaggtgctcaggtccttcagcgtgtgtcgcctctcctcctcgtgaaacagaacc 1056
Db 225 TGGAAAGGTGCTGAGGTCCTTGCAGCTGTGTCGGCTCTCCCTCCGGAACAAAAC 166
Qy 1057 ctccacagcacatcctaccgcgaagaccagcctcagaggtcctcttctgaaacagctgt 1116
Db 165 CTCCACAGCACATCCTACCCGGAAGACCAGCCTCAGAGGTCCTCTCGAACACAGCTGT 106
Qy 1117 ctgtggagagaatgggtgtcttcctcagggactcgtgaagcgtgtgtcctcaggaagac 1176
Db 105 CTGTGGAGAGATGGGTGCTTCTGTCAGGACTGGTGCAGGCTGGTCTCTAAAAAAGGAC 46
Qy 1177 aaactgccagacttgaccgaatgaattttatttttctg 1218
Db 45 CCCCTACCCAGGCTTGAGCCCAATTAATAAATAATTTTGTG 4

RESULT 5
LOCUS AA622758 564 bp mRNA EST 21-OCT-1997
DEFINITION np7604.s1 NCI-CGAP-B2 Homo sapiens cDNA clone IMAGE:113230 3', similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ; mRNA sequence.
ACCESSION AA622758
VERSION AA622758.1 GI:2526634
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 564)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1929 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 407.

FEATURES

source

1. 564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1132230"
/clone_lib="NCI-CGAP-B2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is the normalized version of NCI-CGAP-B2.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."
117 a 169 c 143 g 135 t

BASE COUNT
ORIGIN

Query Match 44.1% Score 538; DB 36; Length 564;

Best Local Similarity 68.9% Pred. No. 1.9e-136;

Matches 552; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 663 gatccaggaccacaaatgaacatgaagcaggaactctctctgacttgggagttcc 722
Db 564 GATCCAGGGACCAAAATGTAACGATGAAGACAGAGAACTCTTCTGACTT-GGGAGTTCC 506
Qy 723 taagcaaaactgggtctccgggacagctgggtcttcataggagcacaacacctcag 782
Db 505 TAGCGAAAAACAATGGCTTCGGGACAGTTGGGTCTTCATAGGAGCCAAAGACCTCAGG 446
Qy 783 ggtaaaagccctttgacagttcttaagaacagccacacacacacacacacacacag 842
Db 445 GGTAAAAGCCCTTTGAGCAGTCTTTAAAGAACAGCCACACAAACAAATACGAGGGA 386
Qy 843 tggccagagctgctgagatgagggctgctgctgctgctgctgctgctgctgctgctgct 902
Db 385 TGCCACAGAGCTGCTGGAGATGGAGTCTGCATGCCCGGAGCCATTTTAGGGTGGCTGT 326
Qy 903 ggtcttctcagccaggggctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 962
Db 325 GGCTCTTCTCAGCCAGGGGCTGGAAGAAAGTCTCCTGCTGACTTAGGAGTCAGAGCCCGG 266
Qy 963 cagggtcctgagagagagcaggggtgctgctgctgctgctgctgctgctgctgctgctgct 1022
Db 265 CAGGGGCTGAGGAGGAGGAGGAGGGGGTCTGCTGGAAAGGTGCTGACAGTCTTTCGACG 206
Qy 1023 ctgtgtcgcgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1082
Db 205 CTGTGTGCGGCTCTCTCTCTCGGAAACAGAACCTCCACAGCAGACATCTCTACCCGGAAG 146
Qy 1083 accagctcagaggggtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1142
Db 145 ACCAGCTCAGAGGGTCTCTCTGGAACACAGCTCTCTGTGGAGAGAAATGGGTGCTTTCGT 86
Qy 1143 cagggtcctgagcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1202
Db 85 CAGGGGACTGTCAGGCTGGTCTGAGGAAAGGACAAACTGCCAGACTTTGAGCCCAATTA 26
Qy 1203 aattttatttttctggt 1220
Db 25 AATTTATTTTGTCTGT 8

RESULT 6

AI245950/c

LOCUS

DEFINITION

QK45B04.x1 NCI-CGAP_C08 Homo sapiens cDNA clone IMAGE:1871887 3', similar to TR:P97805 P97805 EF-7 ; mRNA sequence.

ACCESSION

AI245950

VERSION

AI245950.1 GI:3841347

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 538)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

On Jan 19, 1998 this sequence version replaced gi:2152876.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

Qy	1032	gcctctctctcgtggaacagaaacctccacacagacatctctaccccggaagaccagctc	1091
Db	158	GCCTTCCTCTCCTCGGAACAGAACCCCTCCACACACATCTCTACCCGGAACACAGACTC	99
Qy	1092	agagggtccctctggaaccagctctctgtgagagaatgggtgctttctcagggactg	1151
Db	98	AGAGGGTCTTCTGGAACACAGCTCTCTGTGAGAGAATGGGTGCTTCTCGTCGAGGACTG	39
Qy	1152	ctcagggctggtcctgagggaaggaacaaactgccacag	1189
Db	38	CTGACGGCTGTGCTCTGAGGAAGGACAAACTGCCACAG	1
RESULT	8		
AI348065/c			
LOCUS	AI348065	498 bp	mRNA EST 02-FEB-1999
DEFINITION	qp56409.x1 NCI-CGAP Co8 Homo sapiens cDNA clone IMAGE:1927025	3'	
	similar to TR:P97805 P97805 EF-7	;	mRNA sequence.
ACCESSION	AI348065		
VERSION	AI348065.1	GI:4085271	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 498)		
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	On Feb 6, 1998 this sequence version replaced gi:2848226.		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D.		
	cDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	www-blo.llnl.gov/bbrp/image/image.html		
Insert Length:	841	Std Error:	0.00
Seq primer:	-400P from Gibco		
High quality sequence stop:	459.		
FEATURES	Location/Qualifiers		
source	1..498		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1927025"		
	/clone_lib="NCI-CGAP Co8"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B"		
	/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a		
	modified polylinker; 1st strand cDNA was prepared from		
	colon adenocarcinoma, and was then primed with a Not I -		
	oligo(dT) primer. Double-stranded cDNA was ligated to Eco		
	RI adaptors (Pharmacia), digested with Not I and cloned		
	into the Not I and Eco RI sites of the modified pT73		
	vector. Library is normalized. Library was constructed by		
	Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	98 a 154 c 131 g 115 t		
ORIGIN			
Query Match	40.8%;	Score 498;	DB 45; Length 498;
Best Local Similarity	100.0%;	Pred. No. 1.5e-125;	
Matches 498;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	723	tacgcaaaacactgggcttcctcggaacagctgggtctctcatagagacaaagacctcagg	782
Db	498	TACGCAAAACACTGGGCTTCCTCGGACAGCTGGGTCTTCATAGAGCAAGCAAGCTCAGG	439

```

/clonelib="IMAGE:1013496"
/tissue_type="NCI_CGAP_Co2"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: colon; Vector: Bluescript SK-; Site_1:
Scori; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
CTCAGATTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."
BASE COUNT      125 a 175 c 154 g 140 t
ORIGIN

```

```

Query Match      40.7%; Score 496.8; DB 35; Length 594;
Best Local Similarity 94.7%; Pred. No. 3.4e-125;
Matches 557; Conservative 0; Mismatches 27; Indels 4; Gaps 4;

QY 633 ggtgactggtgctggtgctcctacgacgacgagcaggaacacaaatgaagcagc 692
DB 594 GGTTCACCTGCTGCTGCTGCTTACGAGATTCAGGACCCAAATGAACGATGAAAGC 535
QY 693 agaaactctctgactggtggtggtcctacgcaaaactggtcctcgggacagc 752
DB 534 GGGAAACTCTCTGTCGACCGGGGAGCTCCTACGCAAAACAATGGGCTTCGGACACG 475
QY 753 tgggtcttcattagggccaaagacctcaggggtgtaaaagccctttgagcagttcttaag 812
DB 474 TGGGTCTTCAT-GGAGCCAAAGACCTCAGGGGT-AAAGCCCCCTTTGAGCAATTTTAAAG 417
QY 813 aacagcccgagacacaaataacagggatggtgagcagagctgctgagatgagggctgc 872
DB 416 AAC-GCCCAAGACACAAACAAATACGAGGGATGGCCAGAGCTGCTGAGATGGA-GGCTGC 359
QY 873 atgccccgaagcatttaggtggtgctgctctctcagccagggcctgaagaag 932
DB 358 ATGCCCCCGAAGCCATTTAGGGTGGCTGTGCTTCTCTACGCCAGGGGCTGAAGAAG 299
QY 933 ctctgctgacttaggagtcagagccggcaggggctgagagagagcaggggggtgc 992
DB 298 CTCTCTGCTGACTTAGGAGTCAGAGCCCGCAGGGGCTGAGGAGGAGGAGGGGTGC 239
QY 993 tgcgtggaaggctgctgaggtcctgacgctgtgtcgcgcctcctcctcggaaacag 1052
DB 238 TCGGTGGAAGGTGCTGACAGTCTTGTGACAGCTGTGTGCGGCTCTCTCTCGGAACAG 179
QY 1053 aacccctccacagcacatctaccggaagaccagcctcagaggtctctctggaaccag 1112
DB 178 AACCTTCCACAGCACATCTACCCGGAAGACACGCTCAGAGGGTCTCTCTGGAACAG 119
QY 1113 ctgtctgtgaggaatgggtgcttctgctcagggactgctgacggctggtcctgagaa 1172
DB 118 CTGCTGTGGAGAAATGGGTGCTTCTGTCAGGGACTGCTGACGGCTGGTCTCAGGAA 59
QY 1173 ggaacaactgcccagactgagcccaataattttatttttgcgtggt 1220
DB 58 GGACAAACTGCCAGACTTTGAGCCCAATTTAAATTTTATTTTGTGCTG 11

```

```

RESULT 10
AL01582/c 508 bp mRNA EST 14-DEC-1999
LOCUS to91c11.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2185652 3'
DEFINITION similar to SW:EF07_MOUSE P97805 PROTEIN EF-7, mRNA sequence.
ACCESSION AL01582
VERSION AL01582.1 GI:5367054
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 508)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

```

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
On Mar 10, 1998 this sequence version replaced gi:2948662.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1860 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 352.

FEATURES
source

1. 508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2185652"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/notes="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.69 Kb. Life Technologies catalog #:
11549-011"

BASE COUNT 102 a 158 c 133 g 115 t
ORIGIN

```

Query Match      40.8%; Score 495.4; DB 60; Length 508;
Best Local Similarity 99.6%; Pred. No. 7.7e-125;
Matches 507; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 710 ctgggagttctctacacaaactggcttccgggacagctgggtctctcatagagc 769
DB 508 CTGGGAGTCTTACCCAAAC-ACITGGGCTTCCGGGACAGCTGGGTCTTCATAGGAGC 450
QY 770 caaagacctcaggggtgtaaaagccctttgagcagttctttaaagaacagccacacaaa 829
DB 449 CAAAGACCTCAGGGGTAAAGCCCTTTGAGCAGTCTTAAAGAACAGCCACACACAA 390
QY 830 caaatacagggatggtccagagctgctggagatggagggtgctgcatgccccgaagcatt 889
DB 389 CAAATACGAGGGATGGCAGAGCTGCTGAGATGGAGGGCTGCATGCCCGAAGCCATT 330
QY 890 ttaggtggtgctggtctctctcagccagggcctgaagaagctcctgctgacttagg 949
DB 329 TTAGGGTGGCTGTGGCTCTCTCAGCCAGGGGCTCTGGAAGCTCTGCTGACTTAGG 270
QY 950 agtcagagcccgaggggctgagggagagcagggggtgctgctggtgaaagtctgc 1009
DB 269 AGTCAGAGCCCGCAGGGGCTGAGGAGGAGGAGGAGGGGTGCTGCGTGAAGTCTGCTG 210
QY 1010 aggtccttgacgctgtgtcgccctctctcctcctcctcctcctcctcctcctcctcctc 1069
DB 209 AGGTCTTGTGACGCTGTGTGCGCCTCTCTCTCTCGGAAACAGAACCCCTCCACACACA 150
QY 1070 tctaccggaaacacagcctcagagggctctctctgaaacagctgctgtgagagaat 1129
DB 149 TCTTACCCGGAAGACACGCTCAGAGGGTCTCTCTGGAACACAGCTGTCTGTGGAGAGAT 90
QY 1130 ggggtgcttctcagggactgctgacggctggtcctcctgaggaaggaacactcccgagc 1189
DB 89 GGGGTGCTTCTGTCAGGACTGCTGACGGGTGGTCTCTGAGGAAGGACAAACTGCCAGAC 30
QY 1190 ttgagcccaatttaatttttttttttttttttttttttttttttttttttttttttttt 1218

```

Db	29	TTGAGCCCAATTAAATTTATTTTCTG	1	
RESULT	11			
AA622120/C				
LOCUS		518 bp	mrna	EST
DEFINITION		14-OCT-1997		
		ng5603.s1 NCI-CGAP_Co9 Homo sapiens cDNA clone IMAGE:1147853 3'		
		similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ; mRNA		
		sequence.		
ACCESSION		AA622120		
VERSION		AA622120		
KEYWORDS		AA622120.1 GI:2525996		
SOURCE		EST.		
ORGANISM		human.		
REFERENCE		Homo sapiens		
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
TITLE		Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL		1 (bases 1 to 518)		
COMMENT		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
		Unpublished (1997)		
		On Sep 12, 1996 this sequence version replaced gi:1404591.		
		Contact: Robert Strausberg, Ph.D.		
		Tel: (301) 496-1550		
		Email: Robert.Strausberg@nih.gov		
		Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,		
		M.D., Ph.D.		
		cDNA Library Preparation: M. Bento Soares, Ph.D.		
		cDNA Library Arrayed by: Greg Lennon, Ph.D.		
		DNA Sequencing by: Washington University Genome Sequencing Center		
		Clone distribution: NCI-CGAP clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LLNL at:		
		www-bio.llnl.gov/bbrp/image/image.html		
		Seq primer: -40m13 fwd. ET from Amersham		
		High quality sequence stop: 414.		
FEATURES		Location/Qualifiers		
source		1..518		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:1147853"		
		/clone_lib="NCI-CGAP_Co9"		
		/tissue_type="colon tumor RER+"		
		/lab_host="DH10B"		
		/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a		
		modified polylinker; 1st strand cDNA was prepared from		
		RER+ colon tumor, and was then primed with a Not I -		
		oligo(dT) primer. Double-stranded cDNA was ligated to Eco		
		RI adaptors (Pharmacia), digested with Not I and cloned		
		into the Not I and Eco RI sites of the modified p7T3		
		vector. Library is not normalized. Library was		
		constructed by Bento Soares and M. Fatima Bonaldo		
		(Soares4)." 103 a 158 c 133 g 124 t		
BASE COUNT				
ORIGIN				
Query Match		40.5%;	Score 495;	DB 36; Length 518;
Best Local Similarity		99.8%;	Pred. No. 1e-124;	
Matches		506; Conservative	0; Mismatches	0; Indels
			1; Gaps	1;
QY	715	ggattcctacgc-aaacaactggctccggagcagctgggtcttcagagccaaa	773	
Db	518	GGAGTTCTACGCAAAACAACTGGGCTTCGGGACAGCTGGGTCTTCATAGAGCCAAA	459	
QY	774	gacctcgggttaaaagccctttgagcagttcttaagaacacagccacacacaaa	833	
Db	458	GACCTCAGGGTAAAGCCCTTTGACAGTCTTAAAGACAGCCACACACAAACAA	399	
QY	834	tacgagggatggccagagctgctggagatggagggctgcatgcccccgagccatttag	893	
Db	398	TACGAGGGATGGCCAGAGCTGCTGGAGATGGAGGCTGCATGCCGCCGAAGCATTTAG	339	

BASE COUNT	93 a	149 c	125 g	112 t	
ORIGIN					
Query Match	38.6%; Score 471.6; DB 63; Length 479;				
Best Local Similarity	99.2%; Pred. No. 2.4e-118;				
Matches	474; Conservative	0; Mismatches	4; Indels	0; Gaps	0;
QY	730 acaactgggtccgagcagctgggtctctcattagagccaagacccacaggggtaaaa				789
Db	478 AACAACTGGGCTCCGGGACAGCTGGTCTCTATAGAGGCCAAAGACCTCAGGGGTAAAA				419
QY	790 gccctttgagcgtttttaaaagacagccacacaaacaatacaggggatgccag				849
Db	418 GCCCCTTTGAGCAGTCTTAAAGAACAGCCACACACAAACAAATACGAGGATGCCAG				359
QY	850 agctgtgagatggagggtcgtatcccccaagcatttaggttgcttggtctctt				909
Db	358 AGCTGTGGAGATGGAGGGTGCATGCCCCAGAAAGCCATTTTAGGGTGGCTGTGGCTTTT				299
QY	910 cctcagccaggggctgaagaagctcctgcctgacttagagtcagagcccgaggggc				969
Db	298 CCTCAGCCAGGGCCTTAAGAACTTCTCGCTGACTTAGAGTCAGAGCCCGCAGGGGC				239
QY	970 tgaggaggagcaggggtgctgctggaaggtgctcaggtccttgcaagctgtgtc				1029
Db	238 TGAGGAGGAGGAGCAGGGGTGCTGCTGGAAGGTGCTGCAGGTCCTTGACGCTGTGTGC				179
QY	1030 gcgcctctctcgaagaacagccctccacagacacatctaccgggaagaccagcc				1089
Db	178 GCCTCTCTCTCGGAACAGAACCCCTCCACAGCATCTCTACCCGGAAGACGAGCC				119
QY	1090 tcagaggggtctctggaaccagctgtctgtgagagaatgggtgcttctcagggac				1149
Db	118 TCAGAGGGTCTTCTGSAACCACTGCTGTGGAGAGATGGGGTCTTCTCAGGGAC				59
QY	1150 tgcgtcagcgtgctctgaggaaggacaactgcccagacttgagcccaatttaattt				1207
Db	58 TGTCTAGCGGTGCTCTGTAGGAGGAGCAAACTGCCAGACTTTCAGCCCAATTAATTT				1
RESULT 13					
AA847242/c					
LOCUS	AA847242 464 bp mRNA EST 31-MAR-1998				
DEFINITION	of01b11.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1419837				
ACCESSION	AA847242				
VERSION	AA847242.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 464)				
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
	Unpublished (1997)				
	On Jan 19, 1998 this sequence version replaced gi:2153278.				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				
	Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.				
	cDNA Library Preparation: Stratagene, Inc.				
	cDNA Library Arrayed by: Greg Lennon, Ph.D.				
	DNA Sequencing by: Washington University Genome Sequencing Center				
	Clone distribution: NCI-CGAP clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	www-bio.llnl.gov/bbrp/image/image.html				
	Insert Length: 1265 Std Error: 0.00				
	Seq primer: -40m13 fwd. ET from Amersham				
FEATURES					
Location/Qualifiers					
1..464					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="IMAGE:1419837"					
/clone_lib="NCI_CGAP_Col2"					
/sex="mixed"					
/tissue_type="colon tumor"					
/lab_host="SOLR (kanamycin resistant)"					
/note="organ: colon; Vector: Bluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled colon tumors. 5' adaptor sequence: 5' GAATTCGCGACGAG 3' 3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGT 3' Average insert size: 1.2 kb."					
BASE COUNT	82 a	149 c	126 g	107 t	
ORIGIN					
Query Match	38.0%; Score 464; DB 39; Length 464;				
Best Local Similarity	100.0%; Pred. No. 2.8e-116;				
Matches	464; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	726 gcaaaacaactgggtcccgagcagctgggtctctcattagagccaagacccacaggggt				785
Db	464 GCAAAACAACCTGGGCTCCGGGACAGCTGGTCTCTATAGAGGCCAAAGACCTCAGGGGT				405
QY	786 aaagccctttgagcagcttcttaagaacagccacacaaacaatacaggggatgg				845
Db	404 AAAAGCCCTTTGAGCAGTCTTAAAGAACAGCCACACAAACAAATACGAGGATGG				345
QY	846 ccagagctgctggagatggaggctgcagccccgaagcatttttaggttgctgtggc				905
Db	344 CCAGAGCTGCTGGAGATGGAGGGTGCATGCCCCGAAGCCATTTTAGGGTGGCTGTGGC				285
QY	906 tcttctcagcaggggctggaagaagctcctgcctgacttagtagtcagagcccgag				965
Db	284 TCTTCTCAGCAGCGGGCTCGAAGAGCTCCTGCTGCTTAGGAGTCAGAGCCCGCAG				225
QY	966 gggctgaggagagagcagggggtgctgctggaaggtgctgaggtccttgcacgctg				1025
Db	224 GGGCTGAGGAGGAGGAGCAGGGGTGCTGCTGGAAGGTGCTGCAGGTCCTTGCACGCTG				165
QY	1026 tgcgcgcctctctcctcgaaacagacacccctccacagacacatctaccgggaagacc				1085
Db	164 TGTGCGGCTCTCTCTCTCGGAACACACACCTCCACAGACATCTCCTCCGGAAGACC				105
QY	1086 agcctcagaggggtccttcttgaaaccagctgctgtgagagaatggggtgctttcgtcag				1145
Db	104 AGCCTCAGAGGGTCTTCTGGAACACAGCTGCTGTGGAGAGATGGGGTCTTTCGTCAG				45
QY	1146 ggaactgtacggctgctcctcaggaaggaggaactcctccagac				1189
Db	44 GGACTGTGACGGTGGTCTCTGTGGAAGAGGACAAACTGCCAGAC				1
RESULT 14					
AA622570/c					
LOCUS	AA622570 461 bp mRNA EST 21-OCT-1997				
DEFINITION	np21f04.s1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:1116991 3'				
ACCESSION	AA622570				
VERSION	AA622570.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 461)				
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				

JOURNAL
COMMENT

Unpublished (1997)
On Sep 29, 1997 this sequence version replaced gi:1520561.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 681 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 381.

FEATURES

source

1. .461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1116991"
/clone_lib="NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Mixed germ cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTCTTTTTTTTTTTTTTTT 3' Average insert size: 0.7 kb." Average insert size: 0.7 kb." 106 t

BASE COUNT
ORIGIN

Query Match 37.4%; Score 456.2; DB 36; Length 461;
Best Local Similarity 99.3%; Pred. No. 3.8e-114;
Matches 458; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 729 aacaaactgggtctccggagcgtgggtcttcataggagccaaagacctcagggttaa 788
DB 461 AAACAACCTGGCTTCCGTGACAGCTGGGTCTTCATAGGAGCCAAAGACCTTCAGGGGTAA 402
QY 789 agccctttgagcaggtcttaagaacagccagacacaaatacagaggaatgccca 848
DB 401 AGCCCCCTTGACGAGTCTTAAAGACAGCCCCCAGACACAAATAACGAGGGATGGCCA 342
QY 849 gagctgctggagatggagggtgctgcatgcccccgaagccatttaggggtgctgtgctct 908
DB 341 GAGCTGTGGAGATGGAGGGGTGCTGATGCCCCGGAAGCCATTTAGGGTGGCTGTGGCTCT 282
QY 909 tctcagccaggggctgaagaagctcctgctgacttaggtcagagcccgaggggg 968
DB 281 TCCTCAGCCAGGGGCTGAAGAAGCTCTGCTGACTTAGGAGTTCAGAGCCCGCAGGGG 222
QY 969 ctgaggaggaagacaggggtgctgctggaagtactcaggtccttcacagctatgt 1028
DB 221 CTGAGGAGGAGGACAGGGGTGCTGCTGGAAGGTGCTGAGGTCTCTGCACGCTGTGT 162
QY 1029 cgcgctctctctcgaaacagaacctcccacagacacatctaccgcgaagaccagc 1088
DB 161 CGCGCTCTCTCTCGGAACAGACCCCTCCACAGACACATCTACCCGGAGACCCAGC 102
QY 1089 ctgagaggtctcttgaaaccagctgtctgtggagagaatggggtcttctcagggga 1148
DB 101 CTCAGAGGCTCTTCTGGAACACAGCTGCTGTGGAGAGATGGGGTCTTTTCGTACAGGA 42
QY 1149 ctgctgacgctgtgctctgagaaggaacaaactgccccagac 1189
DB 41 CTGCTGACGGTGTCTGTGAGGAAGGACAAACTGCCCCAGAC 1

RESULT 15
AA52362/cLOCUS
DEFINITION

AA52362 454 bp mRNA EST 05-SEP-1997
nk04h09.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1012577 3' similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1355 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 446.

Location/Qualifiers

1. .454

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1012577"

/clone_lib="NCI_CGAP_Co2"

/tissue_type="tumor"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTCTTTTTTTTTTTT 3' Average insert size: 1.1 kb." CTCGAGTCTTTTTTTTTTTT 3' 106 t

BASE COUNT
ORIGIN

Query Match 36.8%; Score 449.4; DB 35; Length 454;
Best Local Similarity 99.8%; Pred. No. 2.7e-112;
Matches 450; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 768 gccaaagacctcagggttaaaagccctttgagcagctttaaagaacagccagcacaca 827
DB 454 GCCAAGACCTCAGGGGTAAAGCCCTTTGAGCAGTCTTAAAGAACACCCAGACACA 395
QY 828 acaaatagcagggatggccagagctgctggagatggagggctgcatgccccgaagcca 887
DB 394 AACAAATACAGGGATGGCCAGAGCTGCTGGAGATGGAGCGCTGCATGCCCGAAGCCA 335
QY 888 ttttaagggtggctgtgctcttctcagcaggggctgaagaagctctgctgactta 947
DB 334 TTTTAGGGTGGCTGTGGCTCTTCCCTCAGCCAGGGGCTGAAGAAGCTCTCGCTGACTTA 275
QY 948 ggagtcagagcccgaggggtgagagagagagaggggtgctgctggaaggtgct 1007
DB 274 GGAGTCAGAGCCCGGAGGGGCTGAGAGAGGAGGAGGAGGGGTGCTGCGTGGAGGTGT 215
QY 1008 gcaggtccttgacgctgtgtgcgcctctctctcgcgaagacagaccctccacagca 1067
DB 214 GCAGGTCTTGACGCTGTGTGCGCGCTCTCTCTCGGAAACAGAACCTCCACAGCA 155

```

Qy 1068 catctaccgggaagaccagcctcagaggggtccttcttggaaccagctgtctgtggagaga 1127
Db      |||||||
Qy 154  CATCTACCCGGAAGACCGCTCAGAGGTCCTTCTGGAAACAGCTGTCTGTGGAGAGA 95
Db      |||||||
Qy 1128 atgggtgcttctcaggaactgctgacggctggctcctgaggaagacaaactgcccag 1187
Db      |||||||
Qy 94  ATGGGGTGCTTTTCGTACGGGACTGCTGACGGCTGGTGCTCTGAGGAGGACAACTGCCCAG 35
Db      |||||||
Qy 1188 acttgagcccaattaaattttttttgtg 1218
Db      |||||||
Qy 34  ACTTGACCCCAATTAAATTTTATTGCTG 4
Db      |||||||

```

Search completed: May 31, 2000, 17:51:15
Job time: 3432 sec

W P S R E L F H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:04:55 2000; Maspar time 21.50 Seconds
Tabular output not generated. 719.210 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVLRLLALIFAIVTW.....KYEGWPELLEMEGCMPPKPF 223

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.504; Variance 84.613; scale 0.538

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	110	6.5	44	11	088417 HYPOTHETICAL 5.0 KD PR	2.54e-02
2	110	6.5	3190	5	001368 CREB-BINDING PROTEIN H	2.54e-02
3	108	6.4	1001	5	094526 TWO P DOMAIN POTASSIUM	4.90e-02
4	103	6.1	358	5	045691 MOLE5.2 PROTEIN.	2.44e-01
5	101	6.0	253	1	029937 HYPOTHETICAL 30.2 KD P	4.58e-01
6	102	6.0	2285	5	077292 KAKAPO (FRAGMENT).	3.35e-01
7	102	6.0	2396	5	077291 KAKAPO (FRAGMENT).	3.35e-01
8	100	5.9	510	5	096387 INOSINE-5'-MONOPHOSHA	6.25e-01
9	99	5.9	531	6	09XS55 UDP-GLUCURONOSYLTRANSF	8.52e-01
10	99	5.9	532	6	09XS56 UDP-GLUCURONOSYLTRANSF	8.52e-01
11	98	5.8	386	2	082844 XYLOSE REPRESSOR.	1.16e+00
12	97	5.8	396	11	0321K1 HSA BINDING PROTEIN 3.	1.57e+00
13	97	5.8	561	2	052072 PHA SYNTHASE.	1.57e+00
14	97	5.8	1134	1	030078 MOLYBDOTERIN OXIDORED	1.57e+00
15	96	5.7	269	5	031888 COSMID C25H3.	2.12e+00
16	96	5.7	320	5	048887 SIMILARITY TO FREQUENI	2.12e+00
17	96	5.7	389	10	080586 PUTATIVE N-MYRISTOYLTR	2.12e+00
18	95	5.6	124	2	086225 HYPOTHETICAL 14.5 KD P	2.87e+00
19	95	5.6	233	3	012131 N-(5'-PHOSPHORIBOSYL)-	2.87e+00
20	94	5.6	240	2	051829 INNER MEMBRANE PROTEIN	3.86e+00

21	95	5.6	350	5	Q17586 C01H6.6 PROTEIN.	2.87e+00
22	94	5.6	547	5	Q23135 SIMILAR TO GAMMA-GLUTA	3.86e+00
23	94	5.6	556	4	Q93000 CHL1 PROTEIN (FRAGMENT	3.86e+00
24	94	5.6	642	4	Q14777 RETINOBLASTOMA-ASSOCIA	3.86e+00
25	94	5.6	906	4	Q92770 HELICASE.	3.86e+00
26	95	5.6	906	4	Q92998 CHL1 POTENTIAL HELICAS	3.86e+00
27	95	5.6	1018	5	Q17874 F46F6.2 PROTEIN.	2.87e+00
28	92	5.5	178	2	Q48283 BACTERIOPHAGE-LIKE GEN	6.95e+00
29	92	5.5	231	5	Q16239 C13D9.5 PROTEIN.	6.95e+00
30	92	5.5	339	5	Q45064 B0212.4 PROTEIN.	6.95e+00
31	92	5.5	350	5	Q19556 F18E2.4 PROTEIN.	6.95e+00
32	93	5.5	469	2	Q66887 HYPOTHETICAL 54.8 KD P	5.18e+00
33	92	5.5	509	3	Q06137 SIMILAR TO FRUCTOSE-2.	6.95e+00
34	93	5.5	533	6	Q46548 UDP-GLUCURONOSYLTRANSF	5.18e+00
35	93	5.5	533	6	Q46549 UDP-GLUCURONOSYLTRANSF	5.18e+00
36	93	5.5	573	2	Q34989 YVRG PROTEIN.	5.18e+00
37	92	5.5	805	5	Q17235 K10F12.5 PROTEIN.	6.95e+00
38	92	5.5	1865	10	Q81909 T7123.15 PROTEIN.	6.95e+00
39	91	5.4	257	1	Q9YAF9 257AA LONG HYPOTHETICA	9.28e+00
40	91	5.4	805	5	Q93719 F43G9.7 PROTEIN.	9.28e+00
41	91	5.4	833	4	Q75969 PROTEIN KINASE A BINDI	9.28e+00
42	91	5.4	1282	2	Q46348 COLLAGENASE PRECURSOR.	9.28e+00
43	91	5.4	1287	4	Q15468 SIL.	9.28e+00
44	90	5.3	388	2	Q53772 PUTATIVE OXIDOREDUCTAS	1.24e+01
45	90	5.3	463	10	Q42794 ASPARTATE AMINOTRANSFE	1.24e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	44 AA.
ID	088417			
AC	088417			
DT	01-NOV-1998 (Tremblrel. 08, Created)			
DT	01-NOV-1998 (Tremblrel. 08, Last sequence update)			
DT	01-NOV-1998 (Tremblrel. 08, Last annotation update)			
DE	HYPOTHETICAL 5.0 KD PROTEIN (FRAGMENT).			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	CABIN D.E., MCKEE-JOHNSON J.W., MATEJIC L.E., WILTSHIRE T.,			
RA	MTAAVDET A., HUO Y.K., KORENBERG J.R., REEVES R.H.;			
RT	"Physical and Comparative Mapping of Distal Mouse Chromosome 16."			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
KW	EMBL; AF045953; AAC21456.1; -			
DR	Hypothetical protein.			
FT	NON_TER 1			
FT	NON_TER 44			
SQ	SEQUENCE 44 AA; 5010 MW; 296FALDE CRC32;			

Query Match 6.5%; Score 110; DB 11; Length 44;
Best Local Similarity 34.3%; Pred. No. 2.54e-02;
Matches 12; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Db	1	RLKAAKDAIEALGSKRKMKFRSSWVFVAAKGF 35
QY	152	KMDESRKLFSDLGSSYAKQLGFRDSWVFIGAKDL 186
		::: :::: : : ::::: :

RESULT	2	PRELIMINARY;	PRT;	3190 AA.
ID	001368			
AC	001368			
DT	01-JUL-1997 (Tremblrel. 04, Created)			
DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)			
DT	01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE	CREB-BINDING PROTEIN HOMOLOG.			

OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.

```
RX MEDLINE: 97263578.
RA AKIMARU H., CHEN Y., DAI P., HOU D.X., NONAKA M., SMOLIK S.M.,
RA ARMSTRONG S., GOODMAN R.H., ISHII S.;
RT "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
RT signalling.";
RL Nature 386:735-738(1997).
DR EMBL: U88570; AAB53050.1; -.
DR HSSP: P03622; 2IFO.
DR FLYBASE: FBgn0015624; nej.
DR PFAM: PF00439; bromodomain; 1.
DR PFAM: PF00569; Z2; 1.
DR PRINTS: PR00503; BROMODOMAIN.
SQ SEQUENCE 3190 AA; 331879 MW; A44A25BF CRC32;

Query Match 6.5%; Score 110; DB 5; Length 3190;
Best Local Similarity 33.8%; Pred. No. 2.54e-02;
Matches 24; Conservative 14; Mismatches 30; Indels 3; Gaps 3;

Db 181 GGMQVVGTMGMKVMSPMS-TSNNGNMGMAIPGMTIAQNLGNMYLTVNSVGGMGGM 239
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 71 SGAANVVGPTMCFEDRMINSVKNVGRGLN-IALVNGTIGAVLGOKAFDMYSGDVHM-L 128
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 240 VNHKLQOPGGG 250
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 129 VKFLKEIPGA 139

RESULT 3 PRELIMINARY; PRT; 1001 AA.
ID Q94526
AC Q94526;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE TWO P DOMAIN POTASSIUM CHANNEL ORK1.
RN ORK1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RX MEDLINE: 97075152.
RT "ORK1, a potassium-selective leak channel with two pore domains cloned
RT from Drosophila melanogaster by expression in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13256-13261(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A.N., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A.N., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U55321; AAC69250.1; -.
DR FLYBASE: FBgn0017561; ORK1.
KW Ionic channel.
SQ SEQUENCE 1001 AA; 109290 MW; A57C4B04 CRC32;

Query Match 6.4%; Score 108; DB 5; Length 1001;
Best Local Similarity 33.3%; Pred. No. 4.90e-02;
Matches 20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;

Db 97 FFAFTVCSTVGNTISPTT-FAGRMIMTAY-SVIGIPVNGILFAGL-GEYFG-RFFEAIY 152
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 64 YFAFKICSGAA-NVVGPTMCFEDRMINSVKNVGRGLNIALVNGTIGAVLGOKAFD-MY 121
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :

RESULT 4
```

```
ID O45691 PRELIMINARY; PRT; 358 AA.
AC O45691;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MOES-2 PROTEIN.
GN MOES.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA DOBSON R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z93385; CAB07637.1; -.
DR PFAM: PF01018; GTPLOBG; 1.
DR PRINTS: PR00326; GTPLOBG.
SQ SEQUENCE 358 AA; 39356 MW; E44EC56B CRC32;

Query Match 6.1%; Score 103; DB 5; Length 358;
Best Local Similarity 37.3%; Pred. No. 2.44e-01;
Matches 25; Conservative 11; Mismatches 27; Indels 4; Gaps 4;

Db 258 ISFLKHIERCESLWYLDYS-TGS-LTDQVKMLRVELEG-YQKGLGDRASTIVINKIDLS 314
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 129 VKFLKEIPG-GALVIVASYDDPGTKMDESRKLFSDLGSSYAKOLGFRDSWVFIGAKDLR 187
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 315 GKSPREE 321
: ||| | :
QY 188 GKSPFEQ 194

RESULT 5 PRELIMINARY; PRT; 253 AA.
ID O29937
AC O29937;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE HYPOTHETICAL 30.2 KD PROTEIN.
GN AF0307.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VG-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.T., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., RAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
```

RA VENTER J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RL reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997).

DR EMBL: AE001083; AAB90927.1; -.

DR TIGR: AF0307; -.

KW Hypothetical protein.

SQ SEQUENCE 253 AA; 30215 MW; EE14DFE4 CRC32;

Query Match 6.0%; Score 101; DB 1; Length 253;

Best Local Similarity 33.9%; Pred. No. 4.58e-01;

Matches 21; Conservative 15; Mismatches 21; Indels 5; Gaps 4;

Db 190 EYFPEKASVIRTNVQSPDMLEERLKI--IKNEVGKDFYKLMINEEDMVLKEKVVVD 247

Y 63 NYFA-EKICSGAANV-VGPTMCFEDRMINSVKNVYGRGLNIALVN-GTTGAVLGQKAPD 119

Db 248 MY 249

Y 120 MY 121

RESULT 6

ID 077292 PRELIMINARY; PRT; 2285 AA.

AC 077292;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE KAKAPO (FRAGMENT).

GN KAK.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-EMBRYONIC AND IMAGINAL DISC;

RA GREGORY S.L., BROWN N.H.;

RT "Kakapo, a Gene Required for Adhesion Between and Within Cell Layers

RT in Drosophila, Encodes a Large Cytoskeletal Linker Protein Related to

RT Plectin and Dystrophin.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ011925; CAA09870.1; -.

DR HSSP: Q01082; 1AA2.

DR PFAM: PF00307; CH; 2.

DR PFAM: PF00018; SH3; 1.

DR PFAM: PF00435; spectrin; 15.

FT NON_TER 2285 2285

SQ SEQUENCE 2285 AA; 262579 MW; 5F6DAB32 CRC32;

Query Match 6.0%; Score 102; DB 5; Length 2285;

Best Local Similarity 26.5%; Pred. No. 3.35e-01;

Matches 18; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 851 IPGACLLPPP-DOEAIDAEERLKLDRSVALWQKKHLRQNMIFATIRVVGK-WDFD 908

Y 135 IPGGALVLVASYDDPGTKMNDSESRKLF-DLGSSYAKQLGFRDSWVFIGAKDLRGKSPFE 193

Db 909 QFLAMGPE 916

Y 194 QFLKNSPD 201

RESULT 7

ID 077291 PRELIMINARY; PRT; 2396 AA.

AC 077291;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE KAKAPO (FRAGMENT).

GN KAK.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-EMBRYONIC AND IMAGINAL DISC;

RA GREGORY S.L., BROWN N.H.;

RT "Kakapo, a Gene Required for Adhesion Between and Within Cell Layers

RT in Drosophila, Encodes a Large Cytoskeletal Linker Protein Related to

RT Plectin and Dystrophin.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ011924; CAA09869.1; -.

DR HSSP: Q01082; 1AA2.

DR PFAM: PF00307; CH; 2.

DR PFAM: PF00018; SH3; 1.

DR PFAM: PF00435; spectrin; 15.

FT NON_TER 2396 2396

SQ SEQUENCE 2396 AA; 274955 MW; 2D316313 CRC32;

Query Match 6.0%; Score 102; DB 5; Length 2396;

Best Local Similarity 26.5%; Pred. No. 3.35e-01;

Matches 18; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 962 IPGACLLPPP-DOEAIDAEERLKLDRSVALWQKKHLRQNMIFATIRVVGK-WDFD 1019

Y 135 IPGGALVLVASYDDPGTKMNDSESRKLF-DLGSSYAKQLGFRDSWVFIGAKDLRGKSPFE 193

Db 1020 QFLAMGPE 1027

Y 194 QFLKNSPD 201

RESULT 8

ID 096387 PRELIMINARY; PRT; 510 AA.

AC 096387;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

RN [1]

RP SEQUENCE FROM N.A.

RA PROSISE G.L., JAMES A.A., LUECKE H.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF035679; AAD10256.1; -.

DR HSSP: P50097; 1AK5.

DR PROSITE: P500487; IMP_DH_GMP_RED; 1.

SQ SEQUENCE 510 AA; 56148 MW; 196FA02C CRC32;

Query Match 5.9%; Score 100; DB 5; Length 510;

Best Local Similarity 21.2%; Pred. No. 6.25e-01;

Matches 18; Conservative 26; Mismatches 38; Indels 3; Gaps 3;

Db 14 VMSYTVDVII-CMPGVIDFALSDIDLNNMTNITIKTPVISSPMDTVYGHKMSIALALS 72

Y 49 VKRYKGLIKPKCPANYFAFKICS-GAANYVGPTMCFEDRMINSVKNVYGRGLNIAL-VN 106

Db 73 GGLGVTHNNMSIEKQIEEVKKVKRF 97

Y 107 GTTGAVLGQKAFDMYSGDVYMLVKF 131

RESULT 9

ID 09XS55 PRELIMINARY; PRT; 531 AA.

AC 09XS55;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE UDP-GLUCURONOSYLTRANSFERASE.

GN SHEUTG1A6.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae;

```
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAFORK;
RX MEDLINE; 99208485.
RA KOBAYASHI T., TATANO A., YOKOTA H., ONAGA T., WATANABE T., YUASA A.;
RT "Small intestinal UDP-glucuronosyltransferase sheUGT1A07: partial
purification and cDNA cloning from sheep small intestine.";
RL Arch. Biochem. Biophys. 364:143-152(1999).
DR EMBL; AB018477; BAA77457.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 531 AA; 60823 MW; DC5D6014 CRC32;

Query Match 5.9%; Score 99; DB 6; Length 531;
Best Local Similarity 23.3%; Pred. No. 8.52e-01;
Matches 27; Conservative 36; Mismatches 43; Indels 10; Gaps 10;

Db 345 KNTKLKWLKLPNDLLGQPKT-R-AFITH-SGSHGVYEGICNGVPMVMP-L-FGDQMDNA 399
QY 32 KTIRLPRWLASPTKEIQVKYKCGLIKPCPANYFAFK-ICSGAANVVGPTMCFEDRMIMS 90
Db 400 ERMETRAGITLVNLEMSGSDLENALKAVINEKSYKENIMRLSRLHKDRPIEPLD 455
QY 91 PVKNNVGRGLNIALVNGTTGAVLQ-QA-F-DM-YSGDVHVLVKFLKEIPGGALVL 142

RESULT 10
ID Q9XS56 PRELIMINARY; PRT; 532 AA.
AC Q9XS56;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE UDP-GLUCURONOSYLTRANSFERASE.
GN SHEUGT1A07.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAFORK;
RX MEDLINE; 99208485.
RA KOBAYASHI T., TATANO A., YOKOTA H., ONAGA T., WATANABE T., YUASA A.;
RT "Small intestinal UDP-glucuronosyltransferase sheUGT1A07: partial
purification and cDNA cloning from sheep small intestine.";
RL Arch. Biochem. Biophys. 364:143-152(1999).
DR EMBL; AB018478; BAA77458.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 532 AA; 60718 MW; F0E85701 CRC32;

Query Match 5.9%; Score 99; DB 6; Length 532;
Best Local Similarity 23.3%; Pred. No. 8.52e-01;
Matches 27; Conservative 36; Mismatches 43; Indels 10; Gaps 10;

Db 346 KNTKLKWLKLPNDLLGQPKT-R-AFITH-SGSHGVYEGICNGVPMVMP-L-FGDQMDNA 400
QY 32 KTIRLPRWLASPTKEIQVKYKCGLIKPCPANYFAFK-ICSGAANVVGPTMCFEDRMIMS 90
Db 401 ERMETRAGITLVNLEMSGSDLENALKAVINEKSYKENIMRLSRLHKDRPIEPLD 456
QY 91 PVKNNVGRGLNIALVNGTTGAVLQ-QA-F-DM-YSGDVHVLVKFLKEIPGGALVL 142

RESULT 11
ID O82844 PRELIMINARY; PRT; 386 AA.
AC O82844;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE XYLOSE REPRESSOR.
GN XYLR.
```

```
OS Tetragenococcus halophilus (Pediococcus halophilus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Tetragenococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-I-13;
RX MEDLINE; 98320588.
RA TAKEDA Y., TAKASE K., YAMATO I., ABE K.;
RT "Sequencing and characterization of the xyl operon of a gram-positive
bacterium, Tetragenococcus halophila.";
RL Appl. Environ. Microbiol. 64:2513-2519(1998).
DR EMBL; AB009593; BAA31870.1; -.
DR PROSITE; PS01125; ROK; 1.
DR PFAM; PF00480; ROK; 1.
SQ SEQUENCE 386 AA; 43426 MW; B2B31CBB CRC32;

Query Match 5.8%; Score 98; DB 2; Length 386;
Best Local Similarity 29.6%; Pred. No. 1.16e+00;
Matches 21; Conservative 19; Mismatches 29; Indels 2; Gaps 2;

Db 290 EVIQLKNNKALISGVNNIVSMYDPTVINNELYREIPDLINVIEKHILTRNNRVFI 349
QY 124 DVMHLVKEF-LKEIPGALVLVASYDDPCTKMNDESKRFLFSDLGSSYAKQLGFRDSW-VFI 181
Db 350 KNTSLDKTTL 360
QY 182 GAKDLRGKSPF 192

RESULT 12
ID Q921K1 PRELIMINARY; PRT; 396 AA.
AC Q921K1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE HSI BINDING PROTEIN 3.
GN HSI1BP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA TAKEMOTO Y., FURUTA M., SATO M., HASHIMOTO Y.;
RT "Isolation and characterization of a novel HSI SH3 domain binding
protein, HSI1BP3.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ132192; CAAL0600.1; -.
SQ SEQUENCE 396 AA; 43769 MW; ODCA7DDE CRC32;

Query Match 5.8%; Score 97; DB 11; Length 396;
Best Local Similarity 32.6%; Pred. No. 1.57e+00;
Matches 14; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

Db 98 LFGVESDIRERAMFDELLRCVSKDAQLAGSPELLEPLGTRAP 140
QY 179 VFIAGKDLRGKSP-FEQFLKNKSPDNTKNGPELLEMEGCMPP 220

RESULT 13
ID O52072 PRELIMINARY; PRT; 561 AA.
AC O52072;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE PHA SYNTHASE.
GN PHAC.
OS Nocardia corallina.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.
RN [1]
RP SEQUENCE FROM N.A.
RA HALL B., BALDWIN J., RHIE H.G., DENNIS D.;
```

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF01964; AAB94058.1; -
 DR PFAM: PF00561; abhydrolase; 1.
 SQ SEQUENCE 561 AA; 60406 MW; 7A21B91A CRC32;

Query Match 5.8%; Score 97; DB 2; Length 561;
 Best Local Similarity 34.5%; Pred. No. 1.57e+00;
 Matches 19; Conservative 13; Mismatches 21; Indels 2; Gaps 2;

Db 239 VEYSVGRGLQFLSWRNPTAEQDWDFTYAGRVIRAIIDEVREITGSDVDVNLIG 293

Qy 92 VKNVNGRLNALYN-GTTGAVLGQKAFDYMVGDMHLVKFLKEIPGGALV-LVA 144

RESULT 14
 ID O30078 PRELIMINARY; PRT; 1134 AA.

AC O30078;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE MOLYBDOPROTEIN OXIDOREDUCTASE, MOLYBDOPROTEIN BINDING SUBUNIT,

DE PUTATIVE.

GN AF0159.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

OC Archaeoglobus.

RL [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE: 98049343.

RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,

RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,

RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,

RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,

RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,

RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,

RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,

RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,

RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,

RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,

RA VENTER J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997).

DR EMBL: AE001095; AAB91069.1; -

DR TIGR: AF0159; -

DR PFAM: PF01568; Molybdop_binding; 1.

KW Hypothetical protein.

SQ SEQUENCE 1134 AA; 127688 MW; E6FF75AE CRC32;

Query Match 5.8%; Score 97; DB 1; Length 1134;
 Best Local Similarity 27.3%; Pred. No. 1.57e+00;
 Matches 27; Conservative 23; Mismatches 40; Indels 9; Gaps 7;

Db 990 VVAEPTVWKGTAIPVGMGRWADSAVKKPAYFRLNDGVSVAALVS---ELPDGA-SLPS 1045

Qy 88 IMSPVKNNVGRG-LNIALVNGT-TGAVLGQKA-FDMISGDMHLVKFLKEIPGGALVIVA 144

Db 1046 DAVNPVKQDTEKTKRLFTKDRRYDDLGI-DSWRFSG 1083

Qy 145 SYDDPGTKMNDESRR-LFSDLLGSSYAKQLGFRDSWVFIG 182

RESULT 15

ID Q18188 PRELIMINARY; PRT; 269 AA.

AC Q18188;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE COSMID C25H3.

GN C25H3.10.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RL [1]

RN SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSKOUH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS N., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON J., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.;

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA JOHNSON D.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA WATERSTON R.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL: U29535; AAB68788.1; -

SQ SEQUENCE 269 AA; 31750 MW; 6158031C CRC32;

Query Match 5.7%; Score 96; DB 5; Length 269;

Best Local Similarity 37.8%; Pred. No. 2.12e+00;

Matches 17; Conservative 11; Mismatches 14; Indels 2;

Db 148 NOESRNFYHPLSGDHFERK--FRGDWLEIGQNRLRS-SDISTFLKN 189

Qy 154 NDESRKLFSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKN 198

Search completed: Wed May 31 11:05:21 2000

Job time : 26 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Wed May 31 11:04:26 2000;  MasPar time 8.90 Seconds
              762.806 Million cell updates/sec
Tabular output not generated.
```

```
>US-09-186-342-2
Title:
Description: (1-223) from US09186342.pap
Perfect Score: 1686
Sequence: 1 MRVSGVLRLLALIFNIVTW.....KYSGWPELLEMGCMPPKPF 223
```

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database: swiss-prot38
          1:swissprot
```

Statistics: Mean 47.338; Variance 81.778; scale 0.579

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		ID	Description	Pred. No.
		Match	Length			
1	908	53.9	230	1 219_HUMAN	2-19 PROTEIN PRECURSOR	2.14e-179
2	855	50.7	227	1 G786_HUMAN	PROTEIN GS3786.	9.26e-167
3	658	39.0	115	1 O1TL_MOUSE	ONCOPROTEIN INDUCED PR	3.13e-120
4	104	6.2	312	1 DHBK_MOUSE	PUTATIVE STEROID DEHYD	6.70e-02
5	96	5.7	522	1 TSAWR_RCTS	56 KD TYPE-SPECIFIC AN	8.83e-01
6	96	5.7	532	1 TSAKR_RCTS	56 KD TYPE-SPECIFIC AN	8.83e-01
7	95	5.6	124	1 Y48A_HAEN	HYPOTHETICAL PROTEIN H	1.21e+00
8	95	5.6	224	1 TRPF_YEAST	N-(5'-PHOSPHORIBOSYL)A	1.21e+00
9	95	5.6	349	1 FML2_MACMU	N-FORMYL PEPTIDE RECP	1.21e+00
10	95	5.6	885	1 CHS3_EXODE	CHITIN SYNTHASE 3 (EC	1.21e+00
11	92	5.5	215	1 SOMA_TRIVU	SOMATOTROPIN PRECURSOR	3.01e+00
12	91	5.4	166	1 TH12_BOVIN	MITOCHONDRIAL THIOREDO	4.07e+00
13	91	5.4	287	1 DMSC_ECHLO	ANAEROBIC DIMETHYL SUL	4.07e+00
14	91	5.4	32	1 YAM3_SCOLI	HYPOTHETICAL 41.5 KD P	4.07e+00
15	91	5.4	563	1 4CL_ORYSA	4-COUMARATE--COA LIGAS	4.07e+00
16	89	5.3	104	1 ARTA_ECOLI	ARTA PROTEIN.	7.34e+00
17	90	5.3	137	1 YMF5_ECOLI	HYPOTHETICAL 15.4 KD P	5.47e+00
18	89	5.3	159	1 VFPL_FOWP1	PLEFT PROTEIN (FRAGME	7.34e+00
19	89	5.3	204	1 SOMA_SERQU	SOMATOTROPIN PRECURSOR	7.34e+00
20	90	5.3	236	1 MOVK_TMGMV	MOVEMENT PROTEIN (CELL	5.47e+00
21	89	5.3	361	1 SHAK_DROME	SHAKING-B PROTEIN PREC	7.34e+00
22	90	5.3	410	1 CREA_BACSP	CREATINASE (EC 3.5.7.0)	5.47e+00
23	90	5.3	428	1 YMI4_CAEEL	HYPOTHETICAL 47.6 KD P	5.47e+00

[illegible]

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA MEDLINE; 95350630.
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

[illegible]

FT	CHAIN	26	SOMATOPOPEIN.
FT	CHAIN	77	BY SIMILARITY.
FT	DISULFID	188	
FT	DISULFID	213	BY SIMILARITY.
FT	DISULFID	205	

Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
 -!- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
 -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-

SQ SEQUENCE 215 AA; 24353 MW; F241085B4A7352D1 CRC64;

Query Match 5.5%; Score 92; DB 1; Length 215;
Best Local Similarity 32.2%; Pred. No. 3.01e+00;
Matches 28; Conservative 24; Mismatches 28; Indels 7; Gaps 7;

Db 111 LSPVQF-LSRVFNSLVFGTSRDYV-EKLRDLLEG-IQALMQELEDGSGRGLVLTYYD 167
QY 89 MSPVKNNVGRGLNVLNGTTGAVLGQKAFDMYSGDVHVLVFLKEIPG-GALVIVASVD 147
Db 168 KFTNLSRDEALKNYG-LISCFFKDL 193
QY 148 DPC7KM-NDESR-KLFSDLGSSYAKOL 172

RESULT 12

ID TH12 BOVIN STANDARD; PRT; 166 AA.
AC Q95108;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE MITOCHONDRIAL THIREDOXIN PRECURSOR (MT-TRX).
GN TXN2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
[1]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-106 AND 111-166.
RC TISSUE-ADRENAL CORTEX;
RX MEDLINE; 98028383.
RA Watabe S., Hiroi T., Yamamoto Y., Fujioka Y., Hasegawa H., Yago N.,
RA Takahashi S.Y.;
RT "SP-22 is a thioredoxin-dependent peroxide reductase in
RT mitochondria.";
RL Eur. J. Biochem. 249:52-60(1997).
CC -1- FUNCTION: POSSES A DITHIOL-REDUCING ACTIVITY. COOPERATES WITH SP-
CC 22 PROTEIN TO PROTECT GLUTAMINE SYNTHETASE AND OTHER PROTEINS FROM
CC FERROUS ION/DITHIOREITOL-MEDIATED DAMAGE. WITHOUT SP-22, THE
CC PROTEIN HAD NO PROTECTING ACTIVITY.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE THIREDOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87741; BAA13447.1; -
CC PRINTS; PR00421; THIREDOXIN.
CC DR PROSITE; PS00194; THIREDOXIN; 1.
CC DR PFAM; PF00085; thiored; 1.
CC KW Redox-active center; Electron transport; Mitochondrion;
CC TRANSIT peptide.
CC FT TRANSIT 1 59 MITOCHONDRION.
CC FT CHAIN 60 166 MITOCHONDRIAL THIREDOXIN.
CC FT DISULFID 90 93 REDOX-ACTIVE (BY SIMILARITY).
CC SQ SEQUENCE 166 AA; 18416 MW; 5596F7DDA7E34D5F CRC64;

Query Match 5.4%; Score 91; DB 1; Length 166;
Best Local Similarity 19.2%; Pred. No. 4.07e+00;
Matches 15; Conservative 27; Mismatches 31; Indels 5; Gaps 5;

Db 1 MAQRLLRLRLTSLTSIKSPQSRNPASRALKTPQYSPGLTVTPSQARSIVTRVCST 60
QY 17 VTTWVFIRSYMSFKTIR-LPRLWLSPTKEIQVKKYKCG-L-IKPCPA-NYFAFKICSG 72
Db 61 TFNQDGPD-FQDRVYNS 77
QY 73 AANVVGPTMCFEDRMINS 90

RESULT 13
ID DMSO_ECOLI STANDARD; PRT; 287 AA.
AC P18777;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN C (DMSO REDUCTASE ANCHOR
DE SUBUNIT).
DE DMSO.
GN Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / C600;
RX MEDLINE; 89096500.
RA Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT "Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT dimethylsulphoxide reductase of Escherichia coli.";
RL Mol. Microbiol. 2:785-795(1988).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97061202.
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Kimura K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RN RP TOPOLOGY.
RX MEDLINE; 93155163.
RA Weiner J.H., Shaw G., Turner R.J., Trieber C.A.;
RT "The topology of the anchor subunit of dimethyl sulfoxide reductase
RT of Escherichia coli.";
RL J. Biol. Chem. 268:3238-3244(1993).

CC -1- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON
CC VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS. THE C SUBUNIT ANCHORS
CC THE OTHER TWO SUBUNITS TO THE MEMBRANE AND STABILIZE THE CATALYTIC
CC SUBUNITS.
CC -1- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
CC REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSO, A
CC MEMBRANE ANCHOR PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03412; AAA83845.1; -
CC EMBL; AE000191; AAC73982.1; -
CC EMBL; D90727; BAA35628.1; -
CC EMBL; D90728; BAA35631.1; -

```
DR PIR; S03787;
DR ECGENE; EG10234; DMSC.
KW Oxidoreductase; Transmembrane; Inner membrane.
FT DOMAIN 1 9 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 10 32 POTENTIAL.
FT DOMAIN 33 43 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 34 46 POTENTIAL.
FT DOMAIN 44 66 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 67 87 POTENTIAL.
FT DOMAIN 88 107 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 108 112 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 113 134 POTENTIAL.
FT TRANSMEM 135 152 PERIPLASMIC (POTENTIAL).
FT DOMAIN 153 176 POTENTIAL.
FT TRANSMEM 177 182 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 203 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 204 222 POTENTIAL.
FT DOMAIN 223 243 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 244 254 POTENTIAL.
FT DOMAIN 255 281 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 282 287 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 287 AA; 30826 MW; DB6D26ACD2BE0CEB CRC64;

Query Match 5.4%; Score 91; DB 1; Length 287;
Best Local Similarity 23.5%; Pred. No. 4.07e+00;
Matches 8; Conservative 13; Mismatches 12; Indels 1; Gaps 1;

Db 117 LRLIVTWLGVFVMMRVVNSIDVTPTWSIW 150
QY 7 LRL-ALIFAIVTWMFIRSYMSFKTIRLPRW 39

RESULT 14
ID YAM3_SCHPO STANDARD; PRT; 382 AA.
AC Q10058;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 41.5 KD PROTEIN CIF5.03C IN CHROMOSOME I.
OS SPACIF5.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gentles S.; Ogell C.; Churcher C.M.; Barrell B.G.; Rajandream M.A.;
RA Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO YEAST YHR009C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52623; CAA36850.1; -
CC PIR; JU0311; JU0311.
CC HSSP; P08659; ILCI.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC PFAM; PF00501; AMP-binding; 1.
CC KW Ligase; Phenylpropanoid metabolism.
CC SEQUENCE 382 AA; 41501 MW; COAF06F5534C3710 CRC64;

Query Match 5.4%; Score 91; DB 1; Length 382;
Best Local Similarity 28.6%; Pred. No. 4.07e+00;
Matches 22; Conservative 15; Mismatches 35; Indels 5; Gaps 5;

Db 278 LPELSSDTKVDQKCALIKQC-ANHFH-QIIRDSPTKVRQA-CYLFISNATGAPVIGKIG 334
QY 40 LASPTKEIQVKYKGLIKPCPANYFAFKICSGAANNVGTWCF-E-DRMIMSPVKNNVG 97

PIR; S03787;
DR ECGENE; EG10234; DMSC.
KW Oxidoreductase; Transmembrane; Inner membrane.
FT DOMAIN 1 9 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 10 32 POTENTIAL.
FT DOMAIN 33 43 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 34 46 POTENTIAL.
FT DOMAIN 44 66 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 67 87 POTENTIAL.
FT DOMAIN 88 107 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 108 112 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 113 134 POTENTIAL.
FT TRANSMEM 135 152 PERIPLASMIC (POTENTIAL).
FT DOMAIN 153 176 POTENTIAL.
FT TRANSMEM 177 182 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 203 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 204 222 POTENTIAL.
FT DOMAIN 223 243 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 244 254 POTENTIAL.
FT DOMAIN 255 281 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 282 287 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 287 AA; 30826 MW; DB6D26ACD2BE0CEB CRC64;

Query Match 5.4%; Score 91; DB 1; Length 287;
Best Local Similarity 23.5%; Pred. No. 4.07e+00;
Matches 8; Conservative 13; Mismatches 12; Indels 1; Gaps 1;

Db 117 LRLIVTWLGVFVMMRVVNSIDVTPTWSIW 150
QY 7 LRL-ALIFAIVTWMFIRSYMSFKTIRLPRW 39

RESULT 14
ID YAM3_SCHPO STANDARD; PRT; 382 AA.
AC Q10058;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 41.5 KD PROTEIN CIF5.03C IN CHROMOSOME I.
OS SPACIF5.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gentles S.; Ogell C.; Churcher C.M.; Barrell B.G.; Rajandream M.A.;
RA Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO YEAST YHR009C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52623; CAA36850.1; -
CC PIR; JU0311; JU0311.
CC HSSP; P08659; ILCI.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC PFAM; PF00501; AMP-binding; 1.
CC KW Ligase; Phenylpropanoid metabolism.
CC SEQUENCE 382 AA; 41501 MW; COAF06F5534C3710 CRC64;

Query Match 5.4%; Score 91; DB 1; Length 382;
Best Local Similarity 28.6%; Pred. No. 4.07e+00;
Matches 22; Conservative 15; Mismatches 35; Indels 5; Gaps 5;

Db 278 LPELSSDTKVDQKCALIKQC-ANHFH-QIIRDSPTKVRQA-CYLFISNATGAPVIGKIG 334
QY 40 LASPTKEIQVKYKGLIKPCPANYFAFKICSGAANNVGTWCF-E-DRMIMSPVKNNVG 97
```



```

Mfsrcb_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 31 11:03:51 2000; MasPar time 14.02 Seconds
750.250 Million cell updates/sec
Tabular output not generated.

```

```
>US-09-186-342-2
Title:
Description:
Perfect Score: 1686
Sequence: 1 MRYSGVRLRLAIFAIVTW.....KYEGPELLEMEGCMPPKPF 223
```

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:  pir62
           1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 46.426: Variance 93.300: scale 0.498

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	908	53	9	2	I37095	gene 2.19 protein - h	2.76e-154
2	110	6.5	230	2	T13828	CREB-binding protein	6.73e-02
3	108	6.4	1001	2	T13807	potassium channel pro	1.23e-01
4	101	6.0	253	2	C69288	hypothetical protein	9.50e-01
5	102	6.0	2285	2	T13715	kakapo gene protein i	7.13e-01
6	102	6.0	2396	2	T13714	kakapo gene protein -	7.13e-01
7	99	5.9	121	2	S69859	hypothetical protein	1.68e+00
8	97	5.8	383	2	C75153	probable n2,n2-dimeth	2.93e+00
9	97	5.8	1134	2	G69269	molybdopterin-binding	2.93e+00
10	96	5.7	269	2	T15631	hypothetical protein	3.87e+00
11	96	5.7	320	2	T15849	hypothetical protein	3.87e+00
12	96	5.7	389	2	T06697	glycylpeptide N-ter	3.87e+00
13	96	5.7	522	2	B42804	immunodominant type-s	3.87e+00
14	95	5.6	224	1	ISBYN	phosphoribosylanthran	5.09e+00
15	94	5.6	855	2	T10665	hypothetical protein	6.67e+00
16	92	5.5	291	2	T03886	hypothetical protein	1.14e+01
17	93	5.5	469	2	T03057	hypothetical protein	8.74e+00
18	92	5.5	509	2	S51348	hypothetical protein	1.14e+01
19	93	5.5	573	2	B70047	two-component sensor	8.74e+00
20	91	5.4	257	2	F72500	hypothetical protein	1.49e+01
21	91	5.4	287	2	S03787	probable dimethylsulf	1.49e+01
22	91	5.4	493	2	T75078	hypothetical protein	1.49e+01
23	91	5.4	561	1	JU0311	4-coumarate--CoA liga	1.49e+01

ORGANISM	melanogaster)
ENTRY	#formal_name Drosophila melanogaster
DATE	20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSIONS	T13828
REFERENCE	Z17785
#authors	Akamaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.; Goodman, R.H.; Ishii, S.
#journal	Nature (1997) 386:735-738
#title	Drosophila Cbp Is a co-activator of cubitus interruptus in hedgehog signalling.
#accession	T13828
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-3190 ##label AKI
#cross-references	EMBL:U88570; NID:g1916929; PID:g1916930; PIDN:AAB53050.1
GENETICS	
#map_position	X
SUMMARY	#length 3190 #molecular-weight 331879 #checksum 2759
Query Match	6.5%; Score 110; DB 2; Length 3190;
Best Local Similarity	33.8%; Pred. No. 6.73e-02;
Matches	24; Conservative 14; Mismatches 30; Indels 3; Gaps 3;
Db	181 GGMQVVYGTGMVNSPMPS-1SNNNGNMNAIPGMNTIAQGNLGNKVLTVNSVGGMGGM 239
Qy	71 SGAANVVGPTMCFEDRMISPVKNVGRGLN-TALVNGTTGAVLGOKAFDMYSGDVYH-L 128
Db	240 VNHLKQPPGG 250
Qy	129 VKFLKEIPGGA 139
RESULT	3
ENTRY	T13807
TITLE	#type complete
ORGANISM	potassium channel protein - fruit fly (Drosophila melanogaster)
DATE	#formal_name Drosophila melanogaster 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
ACCESSIONS	T13807
REFERENCE	Z17770
#authors	Goldstein, S.A.; Price, L.A.; Rosenthal, D.N.; Pausch, M.H.
#journal	Proc. Natl. Acad. Sci. U.S.A. (1996) 93:13256-13261
#title	ORK1, a potassium-selective leak channel with two pore domains cloned from Drosophila melanogaster by expression in Saccharomyces cerevisiae.
#cross-references	MUID:97075152
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-1001 ##label GOL
#cross-references	EMBL:U55321; NID:g3808067; PID:g3808068; PIDN:AAC69250.1
GENETICS	
#gene	ORK1
#map_position	1
SUMMARY	#length 1001 #molecular-weight 109289 #checksum 4583
Query Match	6.4%; Score 108; DB 2; Length 1001;
Best Local Similarity	33.3%; Pred. No. 1.23e-01;
Matches	20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;
Db	97 FFAPTGVSTGVGNISPTT-FAGRMIMIAV-SVIGIPVNGILFAGL-GRYFG-RTEFAY 152
Qy	64 YFAPKICSGAA-NVVGPTMCFEDRMISPVKNVGRGLNALVNGTTGAVLGOKAFD-MY 121
RESULT	4
ENTRY	C69288
TITLE	#type complete
ORGANISM	hypothetical protein AF0307 - Archaeoglobus fulgidus
ENTRY	#formal_name Archaeoglobus fulgidus

Qy 135 IPGGALVLVASYDDPGTKMDESRKLFSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFE 193
Db 909 QFLAMGPE 916
Qy 194 QFLKNSPD 201
RESULT 6
ENTRY T13714 #type fragment
TITLE kakapo gene protein - fruit fly (Drosophila melanogaster)
ORGANISM (fragment)
DATE #formal_name Drosophila melanogaster
13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
T13714
ACCESSIONS T13714
REFERENCE Z17707
#authors Gregory, S.L.; Brown, N.H.
#journal J. Cell Biol. (1998) 143:1271-1282
#title Kakapo, a gene required for adhesion between and within cell layers in Drosophila, encodes a large cytoskeletal linker protein related.
#accession T13714
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2396 #label GRE
#cross-references EMBL:AJ011924; NID:el332040; PID:el332041;
PIDN:CAA09869.1
GENETICS kak
#note
SUMMARY #length 2396 #checksum 4585
Query Match 6.0%; Score 102; DB 2; Length 2396;
Best Local Similarity 26.5%; Pred. No. 7.13e-01;
Matches 18; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Db 962 IPGACLLPPP-DQEAIDAAERLKRFLDRSVLWQKKHLRLRQNMIFATIRVVKG-WDFD 1019
Qy 135 IPGGALVLVASYDDPGTKMDESRKLFSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFE 193
Db 1020 QFLAMGPE 1027
Qy 194 QFLKNSPD 201
RESULT 7
ENTRY S69859 #type complete
TITLE hypothetical protein YMR075c-a - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
21-Nov-1997
ACCESSIONS S69859
REFERENCE S52814
#authors Pearson, D.; Bowman, S.
#submission submitted to the EMBL Data Library, April 1995
#accession S69859
#molecule_type DNA
#residues 1-121 #label PEA
#cross-references EMBL:248952; MIPS:YMR075c-a
GENETICS
#map_position 13R
SUMMARY #length 121 #molecular-weight 14033 #checksum 4865
Query Match 5.9%; Score 99; DB 2; Length 121;
Best Local Similarity 35.5%; Pred. No. 1.68e+00;
Matches 11; Conservative 12; Mismatches 8; Indels 0; Gaps 0;
Db 79 LLLLLLLLLLWKYKEAMSLTILSLLPWF 109
Qy 9 LLALIFAIVTWMFIRSYMFSMTIRLPW 39
RESULT 8

C75153 #type complete
probleme n2,n2-dimethylguanosine tRNA methyltransferas
PAB2092 - Pyrococcus abyssi (strain Orsay)
#formal_name Pyrococcus abyssi
20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
20-Aug-1999
C75153
ACCESSIONS A75001
REFERENCE
#authors anonymous, Genoscope
#submission submitted to the EMBL Data Library, July 1999
#description Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.
#accession C75153
#status preliminary
#molecule_type DNA
#residues 1-383 #label KAW
#cross-references GB:AJ248284; GB:AL096836; NID:95457730;
PIDN:CAB49306.1; PID:el515201; PID:95457816
#experimental_source strain Orsay
GENETICS
#gene PAB2092
SUMMARY #length 383 #molecular-weight 43218 #checksum 9956
Query Match 5.8%; Score 97; DB 2; Length 383;
Best Local Similarity 21.6%; Pred. No. 2.93e+00;
Matches 16; Conservative 28; Mismatches 26; Indels 4; Gaps 4;
Db 308 QAVKLLKILKDELDPVL-FYDTHG-LGRRLKTEARKI-EEIINEL-RSLGYRASRTHSP 363
Qy 124 DVMHLVKFLKEIPGGALVLVASYDDPGTKMDESRKLFSDLGSSYAKQLGFRDSWVFIGA 183
Db 364 TGVKTDAPYEVFN 377
Qy 184 KDLRGKSPFEQFLK 197
RESULT 9
ENTRY G69269 #type complete
TITLE molybdopterin-binding oxidoreductase homolog - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS G69269
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyriakides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession G69269
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1134 #label KLE
#cross-references GB:AE001095; GB:AE000782; NID:g2689418; PID:g2650483; TIGR:AF0159
SUMMARY #length 1134 #molecular-weight 127688 #checksum 8700
Query Match 5.8%; Score 97; DB 2; Length 1134;
Best Local Similarity 27.3%; Pred. No. 2.93e+00;

```
Matches 27; Conservative 23; Mismatches 40; Indels 9; Gaps 7;

Db 990 VVAAEPTVVKGTIAIPVGMGRWADSAVKKPAYFRNDGSGVAALVS---ELPDGA-SLPS 1045
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 88 IMSPVKNNVGRG-LNTALYNGT-TGAVLGOKA-FDMSGDVHVLKFLKEIFGGALVLVA 144
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 1046 DAVNPVKQLDETAKRILFTKSDRRYYDDLGI-DSWRFSG 1083
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 145 SYDDPGTKMNDESRK-LFSDLGSSYAKQLGFRDSWVFIG 182
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10
ENTRY      T15631      #type complete
TITLE      hypothetical protein C25H3.10 - Caenorhabditis elegans
ORGANISM   #formal_name Caenorhabditis elegans
DATE       20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
ACCESSIONS T15631
REFERENCE  Z18379
#authors   Johnson, D.
#submission submitted to the EMBL Data Library, June 1995
#description The sequence of C. elegans cosmid C25H3.
#accession T15631
##status   preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-269 ##label JOH
##cross-references EMBL:U29535; NID:g868251; PID:g868259;
                  PIDN:AAA68788.1; CESP:C25H3.10
##experimental_source strain Bristol N2
GENETICS
#gene      CESP:C25H3.10
#introns   31/3; 68/3; 140/3; 188/3
SUMMARY    #length 269 #molecular-weight 31750 #checksum 2802

Query Match 5.7%; Score 96; DB 2; Length 269;
Best Local Similarity 37.8%; Pred. No. 3.87e+00;
Matches 17; Conservative 11; Mismatches 14; Indels 3; Gaps 2;

Db 148 NQSRNFYHPLSGDHPRK--FRGDMLEIGNRLRS-SOISTFLKN 189
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 154 NDSRKLFDLGSSYAKQLGFRDSWVFIGAKDLRGKSPPEQLKN 198
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 11
ENTRY      T15849      #type complete
TITLE      hypothetical protein C56C10.9 - Caenorhabditis elegans
ORGANISM   #formal_name Caenorhabditis elegans
DATE       20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
ACCESSIONS T15849
REFERENCE  Z18417
#authors   Fulton, L.
#submission submitted to the EMBL Data Library, June 1995
#description The sequence of C. elegans cosmid C56C10.
#accession T15849
##status   preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-320 ##label FUL
##cross-references EMBL:U29488; NID:g868238; PID:g868247;
                  PIDN:AAA68777.1; CESP:C56C10.9
##experimental_source strain Bristol N2
GENETICS
#gene      CESP:C56C10.9
#introns   73/2; 120/3; 150/1; 202/1; 279/3
SUMMARY    #length 320 #molecular-weight 37556 #checksum 163

Query Match 5.7%; Score 96; DB 2; Length 320;
Best Local Similarity 20.4%; Pred. No. 3.87e+00;
Matches 21; Conservative 32; Mismatches 43; Indels 7; Gaps 6;

Db 201 YDKNNDVVVSQDDFVNGIPGTVDLENLPENNE-KLEKERLRREFNEIDNSDGKATFR 259
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 118 FDMYSDVHVLKFLKEIFGGALVLVASYYDDGTGMNDESR-KLFSD-LGSSYAKQLGFR 175
```

```
Db 260 ELYDYVDQPNERLASKEVNDIMMLTDANNDKLSLEELLERD 301
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 176 DSWFIGAKDLR--GKSPFE-QFLKNSPDTNRYGWPPELLEME 215
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 12
ENTRY      T00697      #type complete
TITLE      glycylopeptide N-tetradecanoyltransferase homolog F6E13.30 -
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
DATE       01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
ACCESSIONS T00697
REFERENCE  Z14198
#authors   Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome II BAC F6E13 genomic
#accession T00697
##status   translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-389 ##label ROU
##cross-references EMBL:AC004005; NID:g3212846; PID:g3212870
##experimental_source cultivar Columbia
GENETICS
#map_position 2
#introns     100/3
#note        F6E13.30
SUMMARY      #length 389 #molecular-weight 44450 #checksum 5324

Query Match 5.7%; Score 96; DB 2; Length 389;
Best Local Similarity 48.5%; Pred. No. 3.87e+00;
Matches 16; Conservative 7; Mismatches 8; Indels 2; Gaps 2;

Db 327 LVNDAL-IVSKQKGFVDYFASDVQMONESEFLKEL 358
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 104 LVNGTTGAVLGQKAFDM-YSGDVHVLKFLKEI 135
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 13
ENTRY      B42804      #type complete
TITLE      immunodominant type-specific antigen tsW56 - Rickettsia
ORGANISM   #formal_name Rickettsia tsutsugamushi
DATE       04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
ACCESSIONS B42804
REFERENCE  A42804
#authors   Ohashi, N.; Nashimoto, H.; Ikeda, H.; Tamura, A.
#journal   J. Biol. Chem. (1992); 267:12728-12735
#title     Diversity of immunodominant 56-kDa type-specific antigen
(TSA) of Rickettsia tsutsugamushi. Sequence and comparative
analyses of the genes encoding TSA homologues from four
antigenic variants.
#cross-references MUID:92316959
#accession B42804
##status   preliminary
##molecule_type DNA
##residues 1-522 ##label OHA
##cross-references GB:M63383; NID:g152512; PID:g152513
##experimental_source strain Kawasaki
##note     sequence extracted from NCBI backbone (NCBIN:108848,
NCBIP:108849)
KEYWORDS   transmembrane protein
SUMMARY    #length 522 #molecular-weight 55971 #checksum 2488

Query Match 5.7%; Score 96; DB 2; Length 522;
Best Local Similarity 31.0%; Pred. No. 3.87e+00;
Matches 18; Conservative 20; Mismatches 17; Indels 3; Gaps 3;
```

```

20-Sep-1999
Tl0665
Zl16533
#authors
M.A.; Barrrell, B.G.; Bancroft, I.; Mewes, H.W.; Mayer,
K.F.X.; Lemcke, K.; Schueller, C.
#submission
submitted to the Protein Sequence Database, June 1999
#accession
Tl0665
#status
preliminary
#molecule_type
DNA
#residues
1-855 #label BEV
#cross-references
EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.30
#experimental_source
cultivar Columbia; BAC clone F6E21
GENETICS
#gene
ATSP:F6E21.30
#map_position
4
#introns
345/1; 756/2
SUMMARY
#length
855 #molecular_weight
95591 #checksum
2864
Query Match 5.6%; Score 94; DB 2; Length 855;
Best Local Similarity 24.0%; Pred. No. 6.67e+00;
Matches 18; Conservative 23; Mismatches 31; Indels 3; Gaps 3;
Db 762 GLDLIQELCFRRIHGNVHSGK-RAMVYSSLUKKANISAKRKQIKLKEINGEKELTFS 820
| : : : | : : | : : | : : | : : | : : | : : | : : | : :
Qy 72 GAANVVGVTMCFEDRTMTSPKYNVGRGLNIA-LVNGTTGAVLGQKA-FDMYSGDVNHLV 129
| : : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 821 EFLKHPSGMEAVINA 835
| : : : | : : | : : | : : | : : | : : | : : | : : | : :
Qy 130 KFLKEIPGQALVLVA 144
| : : : | : : | : : | : : | : : | : : | : : | : : | : :

```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:05:38 2000; MasPar time 6.14 Seconds
Tabular output not generated. 524.106 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVRLALIFAIVTW.....KYGWPELLEMEGCMPPKPF 223

Scoring table: PAM 150
Gap 11

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-Issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 30.975; Variance 139.082; scale 0.223

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	993	58.9	135	4	PCT-US95-0	Sequence 7,	Applicatio	1.52e-80
2	993	58.9	135	1	US-08-469-	Sequence 7,	Applicatio	1.52e-80
3	108	6.4	618	1	US-08-332-	Sequence 2,	Applicatio	1.59e+00
4	102	6.0	993	1	US-08-468-	Sequence 2,	Applicatio	4.40e+00
5	91	5.4	166	2	US-08-775-	Sequence 3,	Applicatio	2.70e+01
6	89	5.3	614	3	US-08-968-	Sequence 13,	Applicatio	3.72e+01
7	89	5.3	680	1	US-07-674-	Sequence 2,	Applicatio	3.72e+01
8	89	5.3	680	2	US-08-436-	Sequence 2,	Applicatio	3.72e+01
9	86	5.1	731	1	US-08-731-	Sequence 2,	Applicatio	6.01e+01
10	85	5.0	76	1	US-08-478-	Sequence 9,	Applicatio	7.03e+01
11	85	5.0	76	1	US-07-881-	Sequence 9,	Applicatio	7.03e+01
12	85	5.0	76	1	US-08-120-	Sequence 9,	Applicatio	7.03e+01
13	84	5.0	119	4	PCT-US95-0	Sequence 8,	Applicatio	8.23e+01
14	84	5.0	119	1	US-08-249-	Sequence 8,	Applicatio	8.23e+01
15	84	5.0	119	2	US-08-886-	Sequence 8,	Applicatio	8.23e+01
16	84	5.0	433	1	US-08-700-	Sequence 20,	Applicatio	8.23e+01
17	85	5.0	671	2	US-08-737-	Sequence 13,	Applicatio	7.03e+01
18	83	4.9	191	1	US-07-885-	Sequence 29,	Applicatio	9.63e+01
19	83	4.9	191	1	US-08-093-	Sequence 3,	Applicatio	9.63e+01
20	83	4.9	193	3	US-08-459-	Sequence 2,	Applicatio	9.63e+01
21	83	4.9	193	2	US-08-383-	Sequence 2,	Applicatio	9.63e+01
22	83	4.9	199	1	US-07-801-	Sequence 4,	Applicatio	9.63e+01
23	83	4.9	217	2	US-09-105-	Sequence 2,	Applicatio	9.63e+01

24	82	4.9	444	3	US-09-221-	Sequence 14,	Applicatio	1.13e+02
25	82	4.9	531	4	PCT-US92-0	Sequence 5,	Applicatio	1.13e+02
26	82	4.9	533	4	PCT-US92-0	Sequence 3,	Applicatio	1.13e+02
27	82	4.9	534	4	PCT-US92-0	Sequence 4,	Applicatio	1.13e+02
28	82	4.9	627	2	US-08-700-	Sequence 6,	Applicatio	1.13e+02
29	82	4.9	627	2	US-08-466-	Sequence 6,	Applicatio	1.13e+02
30	82	4.9	627	3	US-08-467-	Sequence 6,	Applicatio	1.13e+02
31	83	4.9	629	1	US-08-278-	Sequence 6,	Applicatio	9.63e+01
32	83	4.9	629	3	US-08-464-	Sequence 6,	Applicatio	9.63e+01
33	83	4.9	828	1	US-08-261-	Sequence 2,	Applicatio	9.63e+01
34	83	4.9	2104	3	US-08-772-	Sequence 4,	Applicatio	9.63e+01
35	83	4.9	2105	3	US-08-808-	Sequence 4,	Applicatio	9.63e+01
36	83	4.9	2105	2	US-08-772-	Sequence 3,	Applicatio	9.63e+01
37	83	4.9	2105	2	US-08-808-	Sequence 3,	Applicatio	9.63e+01
38	81	4.8	611	2	US-08-677-	Sequence 2,	Applicatio	1.31e+02
39	81	4.8	687	1	US-08-470-	Sequence 10,	Applicatio	1.31e+02
40	81	4.8	687	1	US-08-467-	Sequence 10,	Applicatio	1.31e+02
41	81	4.8	707	4	PCT-US95-1	Sequence 18,	Applicatio	1.31e+02
42	81	4.8	707	1	US-08-528-	Sequence 18,	Applicatio	1.31e+02
43	81	4.8	750	1	US-08-325-	Sequence 2,	Applicatio	1.31e+02
44	81	4.8	750	2	US-08-394-	Sequence 2,	Applicatio	1.31e+02
45	81	4.8	966	2	US-08-868-	Sequence 2,	Applicatio	1.31e+02

ALIGNMENTS

RESULT 1
ID PCT-US95-07289-7 STANDARD; PRT; 135 AA.
XX
AC xxxxxx
DT
XX
XX
Sequence 7, Application PC/TUS9507289
Sequence 7, Application PC/TUS9507289
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07289
FILING DATE: 06-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-265
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 135 AA; 14852 MW; 99029 CN;

CC		NUMBER OF SEQUENCES:	3	
CC		CORRESPONDENCE ADDRESS:		
CC		ADDRESSEE:	Incyte Pharmaceuticals, Inc.	
CC		STREET:	3174 Porter Drive	
CC		CITY:	Palo Alto	
CC		STATE:	CA	
CC		COUNTRY:	USA	
CC		ZIP:	94304	
CC		COMPUTER READABLE FORM:		
CC		MEDIUM TYPE:	Diskette	
CC		COMPUTER:	IBM Compatible	
CC		OPERATING SYSTEM:	DOS	
CC		SOFTWARE:	FastSeq for Windows Version 2.0	
CC		CURRENT APPLICATION DATA:		
CC		APPLICATION NUMBER:	US/08/775,978	
CC		FILING DATE:	To Be Assigned	
CC		PRIOR APPLICATION DATA:		
CC		APPLICATION NUMBER:		
CC		FILING DATE:		
CC		ATTORNEY/AGENT INFORMATION:		
CC		NAME:	Billings, Lucy J.	
CC		REGISTRATION NUMBER:	36,749	
CC		REFERENCE/DOCKET NUMBER:	PF-0176 US	
CC		TELECOMMUNICATION INFORMATION:		
CC		TELEPHONE:	415-855-0555	
CC		TELEFAX:	415-845-4166	
CC		INFORMATION FOR SEQ ID NO:	3:	
CC		SEQUENCE CHARACTERISTICS:		
CC		LENGTH:	166 amino acids	
CC		TYPE:	amino acid	
CC		STRANDEDNESS:	single	
CC		TOPOLOGY:	linear	
CC		IMMEDIATE SOURCE:		
CC		LIBRARY:	GenBank	
CC		CLONE:	1545817	
SQ	SEQ	SEQUENCE	166 AA; 18416 MW; 148927 CN;	
		Query Match	5.4%; Score 91; DB 2; Length 166;	
		Best Local Similarity	19.2%; Pred. No. 2.70e+01;	
		Matches	15; Conservative 27; Mismatches 31; Indels 5; Gaps 5;	
Db	1	MAOELLRLRRLTSLISGKPSQSRWAPVASRALKTPOYSPGYLVTTPSQARSIVTTVRVCS	60	
Qy	17	VTTWMFLRSYMSFKMTIR-LPRWLASPTKEIQVKYKCG-L-IKPCPA-NYEAFKICSG	72	
	:	: :::: : :	: : : : : : :	
Db	61	TFTNIQDGPDPQDRVYNS	77	
	:	: :	: : : :	
Qy	73	AANYVGPTMCDFEDRMIMS	90	
RESULT	6			
ID	US-08-968-563-13	STANDARD;	PRT;	614 AA.
XX	AC	xxxxxx		
XX	XX			
DT	XX			
XX	XX			
DE	Sequence 13,	Application US/08968563		
XX				
CC	Sequence 13,	Application US/08968563		
CC	Patent No. 6013494			
CC	GENERAL INFORMATION:			
CC	APPLICANT:	CHARLES E. NAKAMURA		
CC	APPLICANT:	ANTHONY A. GATENBY		
CC	APPLICANT:	AMY (KUANG-HUA) HSU		
CC	APPLICANT:	RICHARD D. LA REAU		
CC	APPLICANT:	SHARON L. HAYNIE		
CC	APPLICANT:	MARIA DIAZ-TORRES		
CC	APPLICANT:	DONALD E. TRIMBUR		
CC	APPLICANT:	GREGORY M. WHITED		
CC	APPLICANT:	VASANATHA NAGARAJAN		
CC	APPLICANT:	MARK S. PAYNE		
CC	APPLICANT:	STEPHEN K. PICATAGGIO		

CC APPLICANT: RAMSCH V. NAIR
 CC TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
 CC PRODUCTION OF 1,3-PROPANEDIOL
 CC NUMBER OF SEQUENCES: 49
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
 CC STREET: 1007 MARKET STREET
 CC CITY: WILMINGTON
 CC STATE: DELAWARE
 CC COUNTRY: U.S.A.
 CC ZIP: 19898
 CC ADDRESSEE: GENENCOR INTERNATIONAL, INC.
 CC STREET: 4 CAMBRIDGE PLACE
 CC CITY: ROCHESTER
 CC STATE: NEW YORK
 CC COUNTRY: U.S.A.
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: 3.50 INCH DISKETTE
 CC COMPUTER: IBM PC COMPATIBLE
 CC OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
 CC SOFTWARE: MICROSOFT WORD VERSION 7.0A
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/968,563
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 60/030,601
 CC FILING DATE: NOVEMBER 13, 1996
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: FLOYD, LINDA AXAMETHY
 CC REGISTRATION NUMBER: 33,692
 CC REFERENCE/DOCKET NUMBER: CR-9982
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 302-892-8112
 CC TELEFAX: 302-773-0164
 CC INFORMATION FOR SEQ ID NO: 13:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 614 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: unknown
 CC TOPOLOGY: unknown
 CC MOLECULE TYPE: protein
 CC ORIGINAL SOURCE:
 CC ORGANISM: GUT2
 CC SEQUENCE 614 AA: 68811 MW; 2001065 CN;
 Query Match 5.3%; Score 89; DB 3; Length 614;
 Best Local Similarity 35.3%; Pred. No. 3.72e+01;
 Matches 18; Conservative 11; Mismatches 21; Indels 1; Gaps 1;
 Db 49 DAATGRLNVALVEKGFASGTSSKTKMTHGGVRYLEKAFWEFSKAQLDLV 99
 QY 94 NNVGRLNIALVN-GTTGAVLGQKAFDMSYGDVMDHVLKFLKEIPGGALVLV 143
 RESULT 7
 ID US-07-674-287B-2 STANDARD; PRT; 680 AA.
 XX xxxxxx
 AC
 XX
 DT
 XX
 DE Sequence 2, Application US/07674287B
 XX Sequence 2, Application US/07674287B
 CC Patent No. 5414076
 CC GENERAL INFORMATION:
 CC APPLICANT: Bryan Mark O'Hara
 CC TITLE OF INVENTION: Gibbon Ape Leukemia
 CC TITLE OF INVENTION: Virus Receptor
 CC NUMBER OF SEQUENCES: 3
 CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Dr. Karen A. Lowney
 CC ADDRESSEE: American Cyanamid Company
 CC STREET: 1937 West Main Street
 CC CITY: Stamford
 CC STATE: CT
 CC COUNTRY: USA
 CC ZIP: 06904-0060
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy Disk
 CC COMPUTER: IBM PC AT
 CC OPERATING SYSTEM: MS-DOS
 CC SOFTWARE: ASCII converted from IBM DW4
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/674,287B
 CC FILING DATE: 19910325
 CC CLASSIFICATION: 530
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER:
 CC FILING DATE:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Lowney, Karen A., Dr.
 CC REGISTRATION NUMBER: 31,274
 CC REFERENCE/DOCKET NUMBER: 31,104-01
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 203 321 2361
 CC TELEFAX: 203 321 2971
 CC TELEX: 710 474 4059
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 680 Amino Acid Residues
 CC TYPE: AMINO ACID SEQUENCE
 CC STRANDEDNESS: Single
 CC TOPOLOGY: Linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 680 AA; 73675 MW; 2557540 CN;
 Query Match 5.3%; Score 89; DB 1; Length 680;
 Best Local Similarity 28.1%; Pred. No. 3.72e+01;
 Matches 16; Conservative 20; Mismatches 17; Indels 4; Gaps 4;
 Db 81 VSETIRKGLIDVEMYNSTQGLLMAGSVSAMFGSAVWQVLFK-LPISGTHCIVGA 136
 QY 92 VKNVGRGL-NIALVNGTTGAVLGQKAFDMSYGDVMDHVLV-KFLKEIP-GGALVLV 145
 RESULT 8
 ID US-08-436-900A-2 STANDARD; PRT; 680 AA.
 XX xxxxxx
 AC
 XX
 DT
 XX
 DE Sequence 2, Application US/08436900A
 XX Sequence 2, Application US/08436900A
 CC Patent No. 5874264
 CC GENERAL INFORMATION:
 CC APPLICANT: O'Hara, Bryan M.
 CC TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
 CC NUMBER OF SEQUENCES: 4
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: American Home Products
 CC STREET: One Campus Drive
 CC CITY: Parsippany
 CC STATE: New Jersey
 CC COUNTRY: U.S.A.
 CC ZIP: 07054
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/436,900A
CC FILING DATE: 08-MAY-1995
CC CLASSIFICATION: 536
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Barnhard, Elizabeth M.
CC REGISTRATION NUMBER: 31,088
CC REFERENCE/DOCKET NUMBER: 31,104-03
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-683-2158
CC TELEFAX: 201-683-4117
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 680 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 680 AA; 73675 MW; 2557540 CN;

Query Match 5.3%; Score 89; DB 2; Length 680;
Best Local Similarity 28.1%; Pred. No. 3.72e+01;
Matches 16; Conservative 20; Mismatches 17; Indels 4; Gaps 4;

Db 81 VSETIRKGLIDVEMYNSTOGLMAGSVSAMFGSAVNQLVASFLLK-LPISGTHCIVGA 136
QY 92 VKNNVGRGL-NIALVNGTGVAVLGQKAFDMYSGDVNHLV-KFLKEIP-GGALVLVAS 145

RESULT 9
ID US-08-731-716-2 STANDARD; PRT; 731 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08731716
XX
CC Sequence 2, Application US/08731716
CC Patent No. 5789202
CC GENERAL INFORMATION:
CC APPLICANT: Hoskins, JoAnn
CC APPLICANT: Jaskunas, S. Richard
CC APPLICANT: Rocky, Pamela K.
CC APPLICANT: Zhao, Genshi
CC APPLICANT: Rosteck, Paul R. Jr.
CC APPLICANT: No. 5789202ris, Franklin H.
CC TITLE OF INVENTION: Penicillin Binding Protein From
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Eli Lilly and Company
CC STREET: Lilly Corporate Center
CC CITY: Indianapolis
CC STATE: Indiana
CC COUNTRY: U.S.
CC ZIP: 46285
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/731,716
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Webster, Thomas D.
CC REGISTRATION NUMBER: 39,872
CC REFERENCE/DOCKET NUMBER: X-10,887
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 317-276-3334
CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 731 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 731 AA; 80798 MW; 2731494 CN;

Query Match 5.1%; Score 86; DB 1; Length 731;
Best Local Similarity 21.6%; Pred. No. 6.01e+01;
Matches 19; Conservative 24; Mismatches 43; Indels 2; Gaps 2;

Db 472 ATVNDLVGDKAFKFEKGLMKEKVRVLGVAGLVSGVETNPQLMAQAYAFANGLMPEA 531
QY 72 GAANVVGPTCMCFEDRMIMSPVKNNVGRGLNIALVNGTGVAVLG-QKAFDMYSGD-VMLHV 129
Db 532 HFISRIENASQVVIASHKNSQKRVIDKS 559
QY 130 KFLKEIPGGALVLVASVYDDPGTKMNDES 157

RESULT 10
ID US-08-478-675-9 STANDARD; PRT; 76 AA.
XX
AC xxxxxx

XX
DT
XX
DE Sequence 9, Application US/08478675
XX
CC Sequence 9, Application US/08478675
CC Patent No. 5773246
CC GENERAL INFORMATION:
CC APPLICANT: KEENE, JACK D.
CC APPLICANT: KING, PETER H.
CC APPLICANT: LEVINE, TODD
CC TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
CC TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACID
CC NUMBER OF SEQUENCES: 101
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
CC STREET: 1755 Jefferson Davis Highway, Fourth Floor
CC CITY: Arlington
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22202

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,675
CC FILING DATE: 07-JUN-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/120,827
CC FILING DATE: 15-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Oblon, No. 5773246man F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 714-158-0 CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)413-3000
CC TELEFAX: (703)413-2220
CC TELEX: 248855 OPAT UR
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 76 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide

QY	102	IALVNGTTGAVL-GQKAFDMYS--GDVMHLVKFLKEIPGGALV-LVASYD-DPG-TKMD	15
Db	60	N 60	
	:		
QY	156	E 156	
RESULT	12		
ID	US-08-120-827-9	STANDARD; PRT; 76 AA.	
XX	xxxxxx		
DT			
XX			
DE	Sequence 9, Application US/08120827		
CC			
CC	Sequence 9, Application US/08120827		
CC	Patent No. 5525495		
CC	GENERAL INFORMATION:		
CC	APPLICANT: KEENE, JACK D.		
CC	APPLICANT: KING, PETER H.		
CC	APPLICANT: LEVINE, TODD		
CC	TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE		
CC	TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBON		
CC	TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUN		
CC	NUMBER OF SEQUENCES: 101		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,		
CC	ADDRESSEE: P.C.		
CC	STREET: 1755 Jefferson Davis Highway, Fourth Floor		
CC	CITY: Arlington		
CC	STATE: Virginia		
CC	COUNTRY: U.S.A.		
CC	ZIP: 22202		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/120,827		
CC	FILING DATE: 15-SEP-1993		
CC	CLASSIFICATION: 435		
CC	ATTORNEY/AGENT INFORMATION:		
CC	NAME: Oblon, No. 5525495man F.		
CC	REGISTRATION NUMBER: 24,618		
CC	REFERENCE/DOCKET NUMBER: 714-158-0 CIP		
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE: (703)413-3000		
CC	TELEFAX: (703)413-2220		
CC	TELEX: 248855 OPAT UR		
CC	INFORMATION FOR SEQ ID NO: 9:		
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH: 76 amino acids		
CC	TYPE: amino acid		
CC	TOPOLOGY: unknown		
CC	MOLECULE TYPE: peptide		
SQ	SEQUENCE 76 AA; 8481 MW; 29179 CN;		
	Query Match 5.0%; Score 85; DB 1: Length 76;		
	Best Local Similarity 29.5%; Pred.No. 7.03e+01;		
	Matches 18; Conservative 19; Mismatches 17; Indels 7; Gaps		
Db	1	VLVMYGLDQSKMNGDRVFNFCLYGNVER-VKFMKSKPGAAMVEMADGAVDRAITHLNN	59
	:		
QY	102	IALVNGTTGAVL-GQKAFDMYS--GDVMHLVKFLKEIPGGALV-LVASYD-DPG-TKMD	155
Db	60	N 60	
	:		
QY	156	E 156	
RESULT	13		

Fri Jun 2 17:07:45 2000

ID	PCT-US95-06764-8	STANDARD;	PRT;	119 AA.
XX	xxxxxx			
XX				
AC				
XX				
DT				
XX				
XX				
DE	Sequence 8, Application PC/TUS9506764			
XX				
XX				
CC	Sequence 8, Application PC/TUS9506764			
CC	GENERAL INFORMATION:			
CC	APPLICANT: The Regents of the University of			
CC	TITLE OF INVENTION: CLONED leptospira OUTER			
CC	NUMBER OF SEQUENCES: 10			
CC	CORRESPONDENCE ADDRES:			
CC	ADDRESSEE: Fish & Richardson			
CC	STREET: 4225 Executive Square, Suite 1400			
CC	CITY: La Jolla			
CC	STATE: California			
CC	COUNTRY: USA			
CC	ZIP: 92037			

ID XX
 XX AC
 AC AC
 XX XX
 DT DT
 XX XX
 DE DE
 DE DE
 XX XX
 XX XX
 CC Sequence 8, Application PC/TUS9506764
 CC Sequence 8, Application PC/TUS9506764
 CC GENERAL INFORMATION:
 CC APPLICANT: The Regents of the University of California
 CC TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
 CC NUMBER OF SEQUENCES: 10
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Fish & Richardson
 CC STREET: 4225 Executive Square, Suite 1400
 CC City: La Jolla
 CC STATE: California
 CC COUNTRY: USA
 CC ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06764
FILING DATE: 25-MAY-1995
CLASSIFICATION:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06764
FILING DATE: 25-MAY-1995
CLASSIFICATION:

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Haile, Ph.D., Lisa A.,
CC REGISTRATION NUMBER: 38,347
CC REFERENCE/DOCKET NUMBER: FD3602
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 678-5070
CC TELEFAX: (619) 678-5099
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Haile, Ph.D., Lisa A.,
CC REGISTRATION NUMBER: 38,347,
CC REFERENCE/DOCKET NUMBER: FD3602
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 678-5070
CC TELEFAX: (619) 678-5099
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC

CC	LENGTH: 119 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	IMMEDIATE SOURCE:
CC	CLONE: PupA
CC	FEATURE:
CC	NAME/KEY: Protein
CC	LOCATION: 1

LENGTH: 119 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 IMMEDIATE SOURCE:
 CLONE: Pupa
 FEATURE:
 NAME/KEY: Protein
 LOCATION: 1

CC	SEQUENCE	119 AA; 12942 MW; 73190 CN;
SC	Query Match	5.0%; Score 84; DB 4; Length
CC	Best Local Similarity	29.3%; Pred. No. 8.23e+01;
SC	Matches	22; Conservative 21; Mismatches 27; Indels 0
DB	47 EIVRGATGLMTGAGDSAVVN-VIRKLPDVEVGKNY--ELGWKGDAEFTK	: : : : : :
Ov	134 EIPGGALVLLVVDDPTGKTMDENSRKLSSDLGSSYAKQLGRDFSWVF	: : : : : :

[illegible]

Db	104	DKKPRNATVT	LYRD	118
	:	:	:	:
Qy	193	EQ-FLKNSP	DTNKYE	206
RESULT	14			
ID	US-08-249-013-8	STANDARD;		
XX		PRT;	119	AA.
				XX

Db	104	DKKYPRNATVT	LYD	118	
	:	:	:	:	:
QY	193	EQ-FLKNSPD	INKYE	206	
RESULT 14					
ID	US-08-249-013-8	STANDARD;			
XX		PRT;	119	AA.	

AC	xxxxxx
XX	
XX	
DT	
XX	
XX	
DE	Sequence 8, Application US/08249013
XX	
XX	
CC	Sequence 8, Application US/08249013

AC xxxxxx
XX
DT
XX
DE Sequence 8, Application US/08249013
XX
XX Sequence 8, Application US/08249013
CC

Patent No. 5643754
GENERAL INFORMATION:
APPLICANT: Haake, David A.
TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,013
FILING DATE: 25-MAY-1994

Patent No. 5643754
GENERAL INFORMATION:
APPLICANT: haake, David A.
TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,013
FILING DATE: 25-MAY-1994

CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Tumarkin Ph.D., Lisa A.,
CC REGISTRATION NUMBER: P-38,347
CC REFERENCE/DOCKET NUMBER: PD-3602
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 455-5100
CC TELEFAX: (619) 455-5110
CC INFORMATION FOR SEQ ID NO: 8:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin Ph.D., Lisa A.,
REGISTRATION NUMBER: P-38,347,
REFERENCE/DOCKET NUMBER: PD-3602
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 8:

CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 119 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	IMMEDIATE SOURCE:
CC	CLONE: PUPA
CC	FEATURE:

CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 119 amino acids
CC	TYPE: amino acid
CC	STRAINEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	IMMEDIATE SOURCE:
CC	CLONE: Pupa
CC	FEATURE:

```
CC NAME/KEY: Protein
CC LOCATION: 1
CC SEQUENCE 119 AA; 12942 MW; 73190 CN;
Query Match          5.0%; Score 84; DB 1; Length 119;
Best Local Similarity 29.3%; Pred. NO. 8.23e+01;
Matches 22; Conservative 21; Mismatches 27; Indels 5; Gaps 0
Db      47 EIVRGATGLMTGADPSAVVN-VIRKLOPEVGKNY-ELGWKDKGAETKGVDATLNVNINII
       11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
```

CC NAME/KEY: Protein
CC LOCATION: 1
CC SEQUENCE 119 AA; 12942 MW; 73190 CN;

Query Match 5.0% score 84; DB 1; Length 119;
Best Local Similarity 29.3%; Pred. No. 8.23e+01;
Matches 22; Conservative 21; Mismatches 27; Indels 5; Gaps 4;

Dd 47 EIVRGATGLMTAGDPSAVVN-VIRKLDPEWGKNY--ELCWKDGAETKGVDATLNNNIF 103
 || | | | | | | | | | | | | | | : | : | : |
 +

QY	134	EIPCGALVLVASYDDPGTRMNDSEKRLFSDLGSSYAKQLGFRD	SWVF	IGAK-LRGK	SP
Db	104	DKKYPNNAVTLRYD	118		
QY	193	: : : :	206		
RESULT	15				
ID	US-08-886-863-8	STANDARD;	PRT;	119	AA.

QY	134	EIPGALVLVASYDDPCTKKNDESRKLFSDLGSSYAKQLGFRDSWFI	AKAD-LRCKSPF	192
Db	104	DKKYPRNATVTLRYD	118	
	:	:	:	:
QY	193	EQ-FLKNSPDN KYE	206	
RESULT	15			
ID	US-08-886-863-8	STANDARD;	PRT;	119 AA.

XX
 AC
 XX
 DT
 XX
 DE
 XX
 CC
 CC
 CC

xxxxxx

 Sequence 8, Application US/08886863
 Sequence 8, Application US/08886863
 Patent No. 5824321

XX
 AC
 XX
 XX
 DT
 XX
 DE
 XX
 CC
 CC
 CC

xxxxxx

 Sequence 8, Application US/08886863
 Sequence 8, Application US/08886863
 Patent No. 5824321

GENERAL INFORMATION:
 CC APPLICANT: Haake, David A.
 CC TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
 CC NUMBER OF SEQUENCES: 10
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSER: Spensley Horn Jubas & Lubitz
 CC STREET: 1890 Century Park East, Suite 500

CC GENERAL INFORMATION:
CC APPLICANT: Haake, David A.
CC TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Spensley Horn Jubas & Lubitz
CC STREET: 1880 Century Park East, Suite 500
CC

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 17:51:15 ; Search time 799,5 Seconds
(without alignments)
3391.628 Million cell updates/sec

Title: US-09-186-342-8
Perfect score: 669
Sequence: 1 atgmgngtnwsngngngtnt.....gyatgccnccnaarccntty 669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 9714632
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :	EST:*
	1: em_est1:*
	2: em_est2:*
	3: em_est3:*
	4: em_est4:*
	5: em_est5:*
	6: em_est6:*
	7: em_est7:*
	8: em_est8:*
	9: em_est9:*
	10: em_est10:*
	11: em_est11:*
	12: em_est12:*
	13: em_est13:*
	14: em_est14:*
	15: em_est15:*
	16: em_est16:*
	17: em_est17:*
	18: em_est18:*
	19: em_est19:*
	20: gb_est1:*
	21: gb_est2:*
	22: gb_est3:*
	23: gb_est4:*
	24: gb_est5:*
	25: gb_est6:*
	26: gb_est7:*
	27: gb_est8:*
	28: gb_est9:*
	29: gb_est10:*
	30: gb_est11:*
	31: gb_est12:*
	32: gb_est13:*
	33: gb_est14:*
	34: gb_est15:*
	35: gb_est16:*
	36: gb_est17:*
	37: gb_est18:*
	38: gb_est19:*
	39: gb_est20:*
	40: gb_est21:*
	41: gb_est22:*
	42: gb_est23:*
	43: gb_est24:*
	44: gb_est25:*
	45: gb_est26:*
	46: gb_est27:*
	47: gb_est28:*
	48: gb_est29:*
	49: gb_est30:*
	50: gb_est31:*
	51: gb_est32:*
	52: em_est20:*
	53: em_est21:*
	54: em_est22:*
	55: em_est23:*
	56: em_est24:*
	57: em_est25:*
	58: em_est26:*
	59: gb_est33:*
	60: gb_est34:*
	61: gb_est35:*
	62: gb_est36:*
	63: gb_est37:*
	64: gb_est38:*
	65: em_est27:*
	66: em_est28:*
	67: em_est29:*
	68: em_est30:*
	69: gb_est39:*
	70: gb_est40:*
	71: gb_est41:*
	72: gb_est42:*
	73: gb_est43:*
	74: gb_est44:*
	75: em_est31:*
	76: em_est32:*
	77: em_est33:*
	78: em_est34:*
	79: gb_est45:*
	80: gb_est46:*
	81: gb_est47:*
	82: gb_gss1:*
	83: gb_gss2:*
	84: gb_gss3:*
	85: gb_gss4:*
	86: em_gss1:*
	87: em_gss2:*
	88: em_gss3:*
	89: em_gss4:*
	90: gb_gss5:*
	91: gb_gss6:*
	92: gb_gss7:*
	93: gb_gss8:*
	94: gb_gss9:*
	95: em_gss5:*
	96: em_gss6:*
	97: em_gss7:*
	98: em_gss8:*
	99: em_gss9:*
	100: em_gss10:*
	101: em_gss11:*
	102: gb_gss10:*
	103: gb_gss11:*
	104: em_gss12:*
	105: gb_gss12:*
	106: gb_gss13:*
	107: gb_gss14:*
	108: gb_gss15:*
	109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
C 1	224	33.5	645	79	AW300770	AW300770 xk05g02.x
C 2	223.8	33.5	668	63	AW001287	AW001287 wu27e06.x
C 3	216.6	32.4	306	31	AA297147	AA297147 EST112731
C 4	212.6	31.8	778	62	AI883327	AI883327 fc6e0e1.y
C 5	204	30.5	525	80	AW323823	AW323823 uo77a04.y
C 6	203.2	30.4	615	50	AI691072	AI691072 wf22c03.x
C 7	200.8	30.0	646	62	AI936111	AI936111 wo61h03.x
C 8	192.8	28.8	461	36	AA611725	AA611725 wo89f11.r
C 9	189.8	28.4	268	31	AA297176	AA297176 EST112740
C 10	187.6	28.0	417	31	AA298484	AA298484 EST114112
C 11	173.2	25.9	387	33	AA451058	AA451058 vf86h07.r
C 12	161.4	24.1	518	35	AA563933	AA563933 nk23b02.s
C 13	157.6	23.6	410	24	NA1338	NA1338 yw68g05.r1
C 14	157	23.5	570	74	AW175415	AW175415 fi35h03.y
C 15	154.6	23.1	376	23	H26710	H26710 y114b12.r1
C 16	150.6	22.5	564	36	AA622758	AA622758 np76c04.s
C 17	149	22.3	550	81	AA455134	AA455134 zeh12767
C 18	144.2	21.6	264	38	AA793766	AA793766 vr96e04.r
C 19	144	21.5	538	44	AI245950	AI245950 qk45b04.x
C 20	135.8	20.3	364	80	AW369250	AW369250 OV1-BN000
C 21	126.6	18.9	498	45	AI348065	AI348065 qp56d09.x
C 22	124.6	18.6	519	35	AA592348	AA592348 vo77a08.r
C 23	124	18.5	464	39	AA847242	AA847242 of01b11.s
C 24	122.4	18.3	379	80	AW315283	AW315283 12472 MAR
C 25	122	18.2	518	36	AA622120	AA622120 nq56b03.s
C 26	121.6	18.2	508	60	AI801582	AI801582 to91c11.x
C 27	120.4	18.0	461	36	AA622570	AA622570 np21f04.s
C 28	119.4	17.8	479	63	AW001308	AW001308 wu28c10.x
C 29	119.4	17.8	550	64	AW045110	AW045110 um17g12.y
C 30	118.8	17.8	594	35	AA552519	AA552519 nk14e01.s
C 31	118.4	17.7	232	27	AA047528	AA047528 zk73g01.r
C 32	118	17.6	542	79	AW258184	AW258184 ug31e05.y
C 33	117.6	17.6	328	32	AA362591	AA362591 EST72523
C 34	117.6	17.6	448	26	W54133	W54133 md17f04.r1
C 35	111.8	16.7	226	69	AW107522	AW107522 um17g12.x
C 36	109.8	16.4	485	63	AI964291	AI964291 EST269405
C 37	108.4	16.2	244	32	AA379625	AA379625 EST92751
C 38	108	16.1	361	33	AA408987	AA408987 EST03631
C 39	108	16.1	509	34	AA512284	AA512284 vj19g03.r
C 40	103.6	15.5	248	28	AA078382	AA078382 7P02F05 C
C 41	101	15.1	561	34	AA475410	AA475410 vhl7h03.r
C 42	100.6	15.0	710	41	AI038272	AI038272 oy85g06.x
C 43	99.8	14.9	311	80	AA315346	AA315346 12597 MAR
C 44	94	14.1	573	33	AA402132	AA402132 zu55b03.r
C 45	92.6	13.8	367	31	AA285252	AA285252 PMY0756 K

ALIGNMENTS

RESULT 1
 AW300770/c 645 bp mRNA EST 18-JAN-2000
 LOCUS xk05g02.x1 NCI_CGAP_Col9 Homo sapiens cDNA clone IMAGE:2665874 3',
 DEFINITION similar to SW:EF07_MOUSE P97805 PROTEIN EF-7, mRNA sequence.
 ACCESSION AW300770
 VERSION AW300770.1 GI:6710447
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 645)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jul 8, 1999 this sequence version replaced gi:5420873.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Possible reversed clone: polyT not found
 Seq primer: -400P from Gibco
 High quality sequence stop: 403.
 Location/Qualifiers
 source 1. 645
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2665874"
 /clone_lib="NCI_CGAP_Col9"
 /tissue_type="moderately differentiated adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Normalized to Cot 50. Average insert size 1.32kb. Normalized version of NCI_CGAP_Col8. Library constructed by Life Technologies."
 BASE COUNT 137 a 192 c 156 t 1 others
 ORIGIN
 Query Match 33.5%; Score 224; DB 79; Length 645;
 Best Local Similarity 57.5%; Pred. No. 1e-48;
 Matches 181; Conservative 77; Mismatches 57; Indels 0; Gaps 0;
 Qy 355 gayatgtaywsngngaygtnatgcayyytgtnaarttytnaargarathcngngn 414
 ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 645 GACATGTATTCTGAAGATGTATTGCATCATTAATCTTAAAGAAATTCGCGGGGT 586
 ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Qy 415 gcnytngtngtngcngwsntaygayccngngnacnaaratgaaygaygarwsnmg 474
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 585 GCACGTGCTGGTGGCTCTTACGCCGATCCAGGACCAAAATGACGATGAAGCAGA 526
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Qy 475 aarynttywsngayyytngngwsntaygnaarcarytngnttymngaywsntgg 534
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 525 AAACCTCTCTGATCTGGGGAGTCTCTACGCAAAACAACTGGGCTTCCGGGACAGCTGG 466
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Qy 535 gtnttyathgngnaargayyytngngnaarwsnccnttygarcttytnaaraay 594
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 465 GTTTCATAGAGGCCAAAGACCTCAGGGGTAAAGCCCCCTTGAGCAGTTCTTAAAGAAC 406
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Qy 595 wsnccngayacnaayaartayargngtgccngarytntngaratggargntgyatg 654
 ::||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 405 AGCCACACACACAAATACGAGGGATGCCAGAGCTGCTGGAGATGGAGGCTGCATG 346
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Qy 655 ccnccnaarccntty 669
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 345 CCCCGAAGCCATT 331
 ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 RESULT 2
 AW001287/c 668 bp mRNA EST 27-OCT-1999
 LOCUS wu27e06.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2521282 3' similar to SW:EF07_MOUSE P97805 PROTEIN EF-7
 ;contains M5R1.t2 MER22 repetitive element ;, mRNA sequence.
 ACCESSION AW001287
 VERSION AW001287.1 GI:5848203
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 668)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

KEYWORDS

LOCUS	Accession	Organism	Accession	Accession
DEFINITION	wo61h03.x1	NCI_CGAP_Pr22	Homo sapiens	cDNA clone IMAGE1530114

DEFINITION

```

similar to SW:EP07_MOUSE P97805 PROTEIN EF-7 ; , mRNA sequence.
AI936111 GI:5674981
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 646)
AUTHORS       NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT       On May 19, 1998 this sequence version replaced gi:3138334.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 438.
Location/Qualifiers
1..646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2459861"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT    134 a   188 c   163 g   161 t
ORIGIN

Query Match          30.0%; Score 200.8; DB 62; Length 646;
Best Local Similarity 53.2%; Pred. No. 1.4e-42;
Matches 167; Conservative 76; Mismatches 71; Indels 0; Gaps 0;

QY 356 ayatgtaywngngaygtnatgcyytngntnaarttyytnaaargathccnpgnggg 415
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
Db 645 ACATGTACTCAGCGAATGTTATCGCCCTAGTGAGTTCCTAAAGAAGTCGGTAGGGTG 586
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
QY 416 cnyntngtnyngtcnwnshntaygaycngcnagnacnaaratgaaygaygarwsnmgn 475
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
Db 585 CACTGTGTCTGGAGGCTTCATCCGCCGACACAGGACCCTAAATGATCATGAAGCAGTA 526
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
QY 476 arytnttywngayytnngnwnswntaygcnaarcarytngnttymgngaywsntggg 535
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
Db 525 AACTCTTCTGACTTGGGAGTTCATCCATAAACCTGGGCTTCCGGGACAGCTGGG 466
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
QY 536 tnttythagngcnaaargayytnmngngnaarwsncnttytgarcartyyntnaaraayw 595
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
Db 465 TCATTCAATATGAGCCAGACACTCAGGGGTAAAGCCCCCTTGAGCAGTCTCTTAAAGAAC 406
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
QY 596 ncnngayacnaayaartaaygargntngccnrgarytnytngratcgargntgyatgc 655
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |
Db 405 GCCGACACAATCATATACAGGGATGCCACGAGTCTGTGGAGTGGAGGGTGTGCATGC 346
|:||||::: |:|||::: |:|::: |:|::: |:|::: |:|::: |:|::: |:|::: |

```


For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES	source
Location/Qualifiers	
1. .417	
/organism="Homo sapiens"	
/db_xref="ATCC (inhost):175003"	
/db_xref="taxon:9606"	
/clone_lib="Prostate gland I"	
/sex="male"	
/dev_stage="adult, 21 yrs"	
/note="Organ: prostate; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"	
100 a	114 c 106 g 90 t 7 others
BASE COUNT	
ORIGIN	

[illegible]

RESULT	11				
AA451058					
LOCUS		387 bp	mRNA	EST	04-JUN-1997
DEFINITION		vF66h07:rl Soares_mammary_gland_NMMG Mus musculus cDNA clone			
		IMAGE:850717 5' similar to TR:G452270 G452270 2-19 PROTEIN			
		PRECURSOR. ; mRNA sequence.			
ACCESSION		AA451058			
VERSION		AA451058.1	GI:2164728		
KEYWORDS		EST.			
SOURCE		house mouse.			
ORGANISM		Mus musculus			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
		Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1	(bases 1 to 387)			

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394413.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
wustl.com; R.

IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:502869
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 382.

```
FEATURES
source
  i. .387
  Location/Qualifiers
    organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone_image="IMAGE:850717"
    /clone_lib="Soares_mammary_gland_NDMMG"
    /sex="male"
    /tissue.type="mammary gland"
    /dev_stage="4 weeks"
    /lab_host="DH10a"
```

BASE COUNT	109 a	101 c	95 g	82 t
ORIGIN				

Query Match 25.9%; Score 173.2; DB 33; Length 387;
Best Local Similarity 49.5%; Pred. No. 2.1e-35;
Matches 150; Conservative 71; Mismatches 82; Indels 0;

361	QY	taywsngngaydnatgcaaytngtnaarttytnaargarathccngngngngcnytn	420
51	Db	TCCTCCCGACAGCCCCCACTCTCGCTGAACTTCCTTACGGAATCCAGATAGCACCGTG	110
421	QY	gtnyngtngcnwstlaygaycaycngnaccnaaratgaaygagaygawswngnngnaarytn	480
111	Db	GTCTGCTAGCCNCCCTATGATGATGACCTAGGACCCAAATCATGCGACGACAAATCATGACGCGCTT	170

QY	481	TTYSNGAYTNGDGNWSNNTAYGNCARCYTNGGNTTGMNGAYWSNTLGGGNTCT	540
Db	171	TTCTCCACCTGGGAAGTTCCTACGCCAAGCAGCTGGGCTCCGGGACAGCTGGGCTCTTT	230
QY	541	athqngncnaargayytngmgngnaarwsncnttgyarcarttyynaaraywsnccn	600
Db	231	GTAGTGTCTTAAGACCTCAAGATAAAGCCCATATGACGAGTCTTAAAGAACACACCA	290
QY	601	gayacnaayaartaygargngntggccngarytntngaratgargngntgyatgcncncn	660
Db	291	GAGACAAACAAATATGACGGATGGCCGGAGTGTCTAGAGCTCAGGGCTGTGTGCCACGG	350
QY	661	aar	663
Db	351	aa	353

RESULT	12	
AA563933/c		
LOCUS	AA563933	518 bp mRNA EST
DEFINITION	nk23b02.s1 NCI_CGAP_Coll Homo sapiens cDNA clone IMAGE:1014315 similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ;, mRNA sequence.	
ACCESION	AA563933	
VERSION	AA563933.1	GI:2335572
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1	(bases 1 to 518)

AUTHORS	NCI-CGAP
TITLE	http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1290555. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Stratagene, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. cDNA Sequencing: by Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbr/image/image.html

Insert Length: 650 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 452.

FEATURES	source
1. .518	
Location/Qualifiers	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:1014315"	
/clone_lib="NCI_CGAP_Coll1"	
/tissue_type="tumor"	

```

BASE COUNT
94 a 161 c 138 g 125 t
CTCGATGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb. "
GAATTCGGCAGCAG 3' adaptor sequence: 5'
Oligo dt. Multiple colon tumors. 5' adaptor sequence: 5'
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
/lab_host="SOLR (kanamycin resistant)"
class="ef" tumor

```

BASE COUNT
ORIGIN

Query Match 24.1%; Score 161.4; DB 35; Length 518;
Best Local Similarity 58.0%; Pred. NO. 3.3e-32;
Matches 127; Conservative 58; Mismatches 34; Indels 0; Gaps 0;

Qy 451 acnaaratgaaygaygarwsnmgaaryntnttywsngayytnngnwsnwtaygnaar 510

Db 518 ACCAAATGAACGATGAAAGCAGGAACCTTCTCTGACTTGGGAGTCTCTACGCAAAA 459

Qy 511 carytnggnttymngaywsntgggtnttyathggngcnaargayytngmgnggnaarwsn 570

Db 458 CAACTGGGCTTCGGGACAGCTGGGTTCATAGGAGCCAAAGACCTCAGGGTAAAGC 399

Qy 571 ccnttygarcarttyytnaaraaywsnccngayacnaayaartaaygarggntggccngar 630

Db 398 CCTTTGAGCAGTTCTTAAGAACAGCCAGACACAAACAATACGAGGGATGGCCAGAG 339

Qy 631 ytnytngaratggarggntgyatgccnccnaarcentty 669

Db 338 CTGCTGGAGATGGAGGGCTGCATGCCCCGAAGCCATT 300

RESOL 13
N41338

LOCUS	DEFINITION	SYNOPSIS	EST	24-JAN-1990
20000	YW68g05.r1 Soares_placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA clone IMAGE:257432 5', mRNA sequence.	410 bp	EST	

ACCESSION	N41338
VERSION	N41338.1
	GI:1165369

KEYWORDS	EST.	human.
SOURCE		

ORGANISM homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 410)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Eucheria, Firmates, Catalinini, Romilidze, Romio.

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on the Classroom	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Role of the Student in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Impact of the Teacher on the Student	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the Student in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of the Teacher on the Student	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Role of the Student in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Impact of the Teacher on the Student	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the Student in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of the Teacher on the Student	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On May 8, 1995 this sequence version replaced gi:801446.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 331
Source: IMAGE Consortium, LLNL.
This clone is available royalty-free through LLNL ; contact
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7

seq primer: 17
High quality sequence stop: 331.

```

FEATURES
SOURCE
1. 410
location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:3887042"
/db_xref="taxon:9606"
/clone="IMAGE:257432"
/clone_lib="Soares.place"
/dev_stage="two placenta"
/from_9_weeks_post_concep
/lab_host="DH10B (ampic
/notes="Organ: placenta;
modified polylinker; Sit
strand cDNA was primed w
TGTACCAATCTGAAGTGGAGCG
double-stranded cDNA was
adapters (Pharmacia), di
the Not I and Eco RI sit
(Pharmacia). Library con
M. Fatima Bonaldo."
```

	M.Fatima Bonadio.			
	BASE COUNT			
	ORIGIN			
	100 a	107 c	136 g	
			63 t	4 others

Query Match	23.6%	Score 157.6;	DB 24;	Length 410;
Best Local Similarity	48.4%	Pred. No. 2.9e-31;		
Matches 152;	Conservative	69;	Mismatches 93;	Indels 1;
Gaps 1;				

QY 348 rgcnttygatgtaywsngngagytngatgcayytingnaarTTYtynaargathcc 407
:|||:||:||||| : ||||| : ||: ||||| : ||: ||||| : ||: ||| :
Db 14 GGCTTCTGACATGTGGGCCCGAGATGTCACGACCTGTTGAAGTTATT- CGGCCACTGA 72

Db 14 GCCTTCTGACATGTGGCCGGAGATGTCAACGACCTGTTGAAGTTATT-CGGCCACTGA 72

Qy 408 ngngggngcnytngtntnytngtngcnwsntaygaygayccnggnacnaaratgaaygayga 467

Db 73 CGAAGGCACCCCTGGTGTTCGTGGCATCCTACGACGACCCAGCCACCAAGATGAATGAAGA 132

Qy 468 rwsnmgnarytnttywsngayytnggnwsnwsntaygcnaarcarytnggnnttymngna 527

Db 133 GACCAGAAAGCTCTTCAGTGAGCTGGCAGCAGGAACGCCAAGGAGCTGGCCTTCCGGGA 192

QY 528 ywsntqggqnttatyathqqnqcnaaarqayytnmqnqqnaarwsnccnttyqarcarttyt 587

Db 193 CAGCTGGGTGTTGTCTCGGGCCCAAGGGTGTGCAGAACAAAGAGCCCCCTTTGAGCAGCACGT 252

QY 588 naaraaywsnccngayacnaayaartaaygarggntggcngarytnytngratggargg 647

DB 253 GAAGAACAGTAAGCACAGCAACAAGTACGAAGGCTGGCCCGAGGGCGCTGGAGATGGAAGG 312

648 ntgyatgccnccna 661

DB 313 CTTGTAATCCCGCGGA 326

RESULT 14
AW175415
LOCUS

AW175415	LOCUS	AW175415	570 bp	mRNA	EST	16-NOV-1999

Query Match 23.1%; Score 154.6; DB 23; Length 376;
Best Local Similarity 46.4%; Pred. No. 1.7e-30;
Matches 166; Conservative 75; Mismatches 115; Indels 2;

```

Qy 248 tygargaymgnatgathatqwsnccngntnaaraayaaygtngngmgnngnytnaayathg 307
Db 3 TCGAGGACACAGATGCTGATAGGAGCCTCAAGGACACACGTGGGCCCGGNTGAAACATCG 62

Qy 308 cnytngtnaayggnacnagngngcngntnytingnncaraargcnytygayatgtaywsng 367
Db 63 CCCTGGTGAACGGGGTCAAGCGCGG-ANTCATCGAGGCCCGGGCTTTTGACATGTNGGCCG 121

Qy 368 gngaaygtnatgcayytngtnaarttytnaargarathccngngngngcnytngtntng 427
Db 122 GAGATGTCAACGACCTGTTGAAGTTTATTCGGCCACTGCACAGAGGCACCCCTNGTGTTCG 181

Qy 428 tngcnwsntaygavgyccnggnacnaaratgaaygaygarwsnmgnaarytnttywsng 487
Db 182 TGGCATCTACGAGCGGCCAGCCACCAAGATGATGAGAGACCCAGAAAGCTCTTCAGTG 241

Qy 488 ayytnggnwnsntaygcnaarcarvtyngnttymgngaywsntgggtnttyathggng 547
Db 242 AGCTGGGCAGCAGGAGCCCAAGGAGCTGGCTTCGGGGACAC-TNGGTGTTTTTCGGGG 300

Qy 548 cnaargayytmgngngnaarwsnccnttygarcarttytnaaraaywsnccngayac 605
Db 301 CCAAGGGTGTNCAGAACAGAGGCCCTTTNAGCAGCACGTGAAGACAGTTAAGGCAC 358

```

Search completed: May 31, 2000, 17:51:19
Job time: 3436 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 19:14:15 ; Search time 57.43 Seconds
(without alignments)
1514.191 Million cell updates/sec

Title: US-09-186-342-8

Perfect score: 669

Sequence: 1 atgmngntnswngngntny.....gyatgcnccnaarcontty 669

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/prodata/2/ina/6_COMB.seq:*
6: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	294.6	44.0	548	2	US-08-469-667-6
2	294.6	44.0	548	6	PCT-US95-07289-6
C 3	48.6	7.3	7218	1	US-08-232-463-14
C 4	39.4	5.9	1215	4	US-09-092-770-8
5	35.8	5.4	1058	2	US-08-608-241-3
6	35.8	5.4	1058	3	US-08-922-182-3
7	35.8	5.4	1058	3	US-08-919-953-3
8	35.4	5.3	2262	4	US-08-674-887A-5
C 9	35.2	5.3	774	6	PCT-US91-05766-1
10	35.2	5.3	1564	4	US-08-852-481-1
11	35.2	5.3	3101	4	US-08-868-786-1
C 12	35	5.2	2855	4	US-08-852-153-1
C 13	35	5.2	3255	4	US-08-852-153-5
C 14	35	5.2	3327	4	US-08-852-153-3
15	34.4	5.1	579	2	US-08-190-560-3
16	34.4	5.1	579	2	US-08-469-277-3
17	34.4	5.1	579	3	US-08-468-946-3
18	34.4	5.1	579	4	US-08-468-942-3
C 19	34.4	5.1	736	3	US-08-540-406-1
C 20	34.4	5.1	736	5	US-08-656-055-1
C 21	34.4	5.1	736	6	PCT-US95-13233-1
C 22	34.4	5.1	1212	4	US-09-092-770-18
23	34.2	5.1	9421	3	US-08-370-319C-2
24	34.2	5.1	246240	3	US-08-724-394A-20
25	34.2	5.1	246240	3	US-08-724-394A-21
26	34.2	5.1	246240	3	US-08-724-394A-22
27	33.6	5.0	303	2	US-08-190-560-1

28 33.6 5.0 303 2 US-08-469-277-1 Sequence 1, Appli
29 33.6 5.0 303 3 US-08-468-946-1 Sequence 1, Appli
30 33.6 5.0 303 4 US-08-488-942-1 Sequence 1, Appli
31 33.6 5.0 1269 2 US-08-808-641-2 Sequence 2, Appli
32 33.6 5.0 1269 4 US-09-064-839-2 Sequence 2, Appli
33 33.6 5.0 1269 5 US-09-351-438-2 Sequence 2, Appli
C 34 33.6 5.0 15225 4 US-08-892-403A-2 Sequence 2, Appli
35 33 4.9 366 4 US-08-853-659A-11 Sequence 11, Appli
36 33 4.9 2260 1 US-07-794-393-3 Sequence 3, Appli
37 33 4.9 2260 1 US-08-001-711-3 Sequence 3, Appli
38 33 4.9 15512 4 US-08-853-659A-5 Sequence 5, Appli
C 39 33 4.9 15512 4 US-08-853-659A-8 Sequence 8, Appli
40 33 4.9 15512 4 US-08-853-659A-63 Sequence 63, Appli
C 41 33 4.9 15512 4 US-08-853-659A-66 Sequence 66, Appli
42 33 4.9 24701 4 US-08-853-659A-2 Sequence 2, Appli
C 43 33 4.9 24701 4 US-08-853-659A-3 Sequence 3, Appli
44 33 4.9 24701 4 US-08-853-659A-60 Sequence 60, Appli
C 45 33 4.9 24701 4 US-08-853-659A-61 Sequence 61, Appli

ALIGNMENTS

RESULT 1
US-08-469-667-6
; Sequence 6, Application US/08469667
; Patent No. 5733748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat.peptide
; LOCATION: 1..405
US-08-469-667-6

Query Match 44.0%; Score 294.6; DB 2; Length 548;


```
;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Rhodobacter sphaeroides
; STRAIN: 2.4.1
; IMMEDIATE SOURCE:
; CLONE: pUI8017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 221..895
;
US-08-608-241-3

Query Match          5.4%; Score 35.8; DB 2; Length 1058;
Best Local Similarity 27.3%; Pred. No. 0.36;
Matches 93; Conservative 47; Mismatches 201; Indels 0; Gaps 0;

QY 102 hmgnytncnmngtgytngcnwsnccnaaargathcargtnaaraartayaartg 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 CTTTGGCGGGTTATCGCGTCAGCGGGATGTGCTAGCCTCGTGAAGACGATATGGGAC 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 yggnytnathaaarccntgycngcnaaytaytgyctnttyaarathtgywsngngcngc 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 CGCATGATGCAACCTTCCCGCCGCTCCAGTCCGCGCTGATCGTCGACGATCCGGT 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 naaytngtngnccnaccnatgtgytgygargaymgnatgathatgwsnccngtnaaraa 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 CTTCTGGGATCTGAACCTGCGGACGTGAACGGCTCGACGGCCCTCATCCGCTCGAAGGC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 yaaytngnmngngnytnaayathgcnyngtngtnaayggnacnagngngcngtntngg 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 CGCGACCGGCTCGAGACCGCTGCGCGGCTGACCTGCGACCGCGCTTCGATGTGGT 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 ncaraargcnttygayatgtaywsngngngaygtnatgcayytngtnaarttytnaarga 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 GGTGCTGGATCTGAACCTGCGGACGTGAACGGCTCGACGGCCCTCATCCGCTCGAAGGC 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 rathcngngngngcnyntngtntngtngcnwsntaygayg 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GAGCCCGGGGGTGGCGGTGGTGGTGTCTCTTCCCTCG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-922-182-3
; Sequence 3, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Wittruh, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Rhodobacter sphaeroides
; STRAIN: 2.4.1
; IMMEDIATE SOURCE:
; CLONE: pUI8017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 221..895
;
US-08-922-182-3

Query Match          5.4%; Score 35.8; DB 3; Length 1058;
Best Local Similarity 27.3%; Pred. No. 0.36;
Matches 93; Conservative 47; Mismatches 201; Indels 0; Gaps 0;

QY 102 hmgnytncnmngtgytngcnwsnccnaaargathcargtnaaraartayaartg 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 CTTTGGCGGGTTATCGCGTCAGCGGGATGTGCTAGCCTCGTGAAGACGATATGGGAC 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 yggnytnathaaarccntgycngcnaaytaytgyctnttyaarathtgywsngngcngc 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 CGCATGATGCAACCTTCCCGCCGCTCCAGTCCGCGCTGATCGTCGACGATCCGGT 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 naaytngtngnccnaccnatgtgytgygargaymgnatgathatgwsnccngtnaaraa 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 CTTCTGGGATCTGAACCTGCGGACGTGAACGGCTCGACGGCCCTCATCCGCTCGAAGGC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 yaaytngnmngngnytnaayathgcnyngtngtnaayggnacnagngngcngtntngg 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 CGCGACCGGCTCGAGACCGCTGCGCGGCTGACCTGCGACCGCGCTTCGATGTGGT 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 ncaraargcnttygayatgtaywsngngngaygtnatgcayytngtnaarttytnaarga 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 GGTGCTGGATCTGAACCTGCGGACGTGAACGGCTCGACGGCCCTCATCCGCTCGAAGGC 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 rathcngngngngcnyntngtntngtngcnwsntaygayg 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GAGCCCGGGGGTGGCGGTGGTGGTGTCTCTTCCCTCG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-919-953-3
; Sequence 3, Application US/08919953
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
```

```
> APPLICANT: Withuhn, Vernon
> TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
> TYPE OF INVENTION: SENSING AND REMEDIATION
> NUMBER OF SEQUENCES: 7
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Quarles & Brady
> STREET: 1 South Pinckney Street
> CITY: Madison
> STATE: WI
> COUNTRY: US
> ZIP: 53703
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.30
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/919,953
> FILING DATE:
> CLASSIFICATION: 435
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/608,241
> FILING DATE:
> ATTORNEY/AGENT INFORMATION:
> NAME: Seay, Nicholas J
> REGISTRATION NUMBER: 27,386
> REFERENCE/DOCKET NUMBER: 960296.93511
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 608-251-5000
> TELEFAX: 608-251-9166
> INFORMATION FOR SEQ ID NO: 3:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1058 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: linear
> MOLECULE TYPE: DNA (genomic)
> ORIGINAL SOURCE:
> ORGANISM: Rhodobacter sphaeroides
> STRAIN: 2.4.1
> IMMEDIATE SOURCE:
> CLONE: pui8017
> FEATURE:
> NAME/KEY: CDS
> LOCATION: 221..895
```

[illegible]

```

Db      451 GAGCCCGGGGCGGTGCCTTCTTCCTCG 491
          :|:||||| | | || |||||::|::|:|
RESULT      8
US-08-674-887A-5
; Sequence 5, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2262 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...2259
US-08-674-887A-5

```

	Query Match	5.38;	Score 35.4;	DB 4;	Length 2262;
	Best Local Similarity	41.3%;	Pred.	No. 0.94;	
Matches	43; Conservative	17; Mismatches	44; Indels	0; Gaps	0;
Qy	541 athgngcnaargayytgmngmgnaarwsnccnttygarcarttynytnaarawywnccn	600 : : : :: :	:	: : :	-
Dd	91 ATAAGGAGACGTATCAGCGGTAAATGTCCTGCATGCACGGTGGTAACACTCGACC	150			
Qy	601 gayacnaayaartaygargngttggcgngaryntyntngaratgga	644 ::: :: :: :::	:	:: :: ::	
Dd	151 GGTACTCCCAACAAGAATTGTTGCCGGAAGGGTTGAACCTGGA	194			

```

RESULT          9
PCT-US91-05766-1/c
: Sequence 1, Application PC/TUS9105766
: GENERAL INFORMATION:
: APPLICANT: Rosenblum, Michael
: APPLICANT: Kohr, William Jack
: APPLICANT: Aggarwal, Bharat
: TITLE OF INVENTION: Protein Structure of the Plant Toxin
:

```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,560
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 7879ZY
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 579 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-190-560-3

```

```

Query Match      5.1%; Score 34.4; DB 2; Length 579;
Best Local Similarity 27.8%; Pred. No. 0.57;
Matches 72; Conservative 40; Mismatches 147; Indels 0; Gaps 0;

Qy 350 cnttygavatgtaywsgngngaygtatgcaytytngttnaarttyytnaaragarathccng 409
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 CTTCCACAGTACTCGGCAAGAGGGGTGACAGTTCAAGCTCAACAGTCAGAGTCAAGACTAA 238
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 410 gngngcnyntngtntngtngcnwsntaygayccnggnacnaaratgaaygaygarw 469
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 239 AGGAGCTCTGACCCGGGAGCTGCCAGCTTCTTGGGGAAAAGGACAGATGAAGCTGCTT 298
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 470 snmgnaaryntnttywsgngaytyngnwnsnwsntaygcnarcarrytngnttymgngayw 529
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 299 TCCAGAGCTGATGAGCAACTTGGACACACAGGAGGACACAGGAGGTGGACTTCCAAGAGT 358
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 530 sntgggtnttythgngngnaaragayttnmgngngnaarwsnccnttygarcarttyytna 589
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 359 ACTGTGTCTTCTGTCTGTCATCGCCATGATGTGTAACGAATTCCTTTGAAGGCTTCCAG 418
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 590 araaywscncngayacnaa 608
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 419 ATAAGCAGCCCGAGGAGNA 437
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 31, 2000, 19:14:25
Job time: 6263 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 19:03:54 ; Search time 78.65 Seconds
(without alignments)
2128.144 Million cell updates/sec

Title: US-09-186-342-8

Perfect score: 669

Sequence: 1 atgmngntnwsngngngntny.....gyatgcncnnaarccntty 669

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	485.8	72.6	1299	1 V63516	Consensus sequence
2	447	66.8	955	1 V63515	Full length sequen
3	294.6	44.0	548	1 T45883	Human colon specif
4	294.6	44.0	548	1 V16671	Polynucleotide seq
5	193.6	28.9	263	1 V63509	Clone 1281865 EST
6	181.4	27.1	251	1 V63511	Clone 1735382 EST
7	161.8	24.2	228	1 V63510	Clone 1807758 EST
8	59.2	8.8	287	1 V63508	Clone 3210629 EST
9	40.4	6.0	187	1 V63512	Clone 2128334 EST
10	39	5.8	3830	1 N70938	DNA fragment contg
11	37.8	5.7	704	1 V74910	Staphylococcus aur
12	37	5.5	1341	1 V69614	Human secreted pro
13	36.8	5.5	1633	1 T08144	Xylanase xynB gene
14	36.6	5.5	987	1 V31480	Human interleukin-
15	36.6	5.5	3934	1 T30309	Human SOX-9 cDNA.
16	36.4	5.4	8536	1 V74589	Staphylococcus aur
17	36.4	5.4	110000	1 V21209_09	Continuation (10 o
18	35.8	5.4	1058	1 V62126	R. sphaeroides Adh
19	35.8	5.4	1058	1 V64113	Rhodobacter beta ge
20	35.6	5.3	1836	1 N20057	Interferon-beta ge
21	35.6	5.3	4544	1 V74828	Staphylococcus aur
22	35.4	5.3	1836	1 T29734	Brevibacterium fla
23	35.4	5.3	2262	1 V06554	Alcaligenes (Deley
24	35.2	5.3	3101	1 V32918	Solanum tuberosum
25	35	5.2	420	1 V77877	Staphylococcus aur
26	35	5.2	1442	1 T83899	DNA encoding a Sta
27	35	5.2	1442	1 V53367	Human sex comb on
28	35	5.2	2855	1 V00011	Human sex comb on
29	35	5.2	2855	1 T97297	Human sex comb on
30	35	5.2	3255	1 T97299	Human sex comb on
31	35	5.2	3255	1 V00013	Human sex comb on
32	35	5.2	3327	1 T97298	Human sex comb on
33	35	5.2	3327	1 V00012	Human sex comb on
34	35	5.2	14078	1 V74502	Staphylococcus aur

35	34.8	5.2	488	1 T20500	Human gene signatu
36	34.8	5.2	600	1 V86185	EST clone J635. Ne
37	34.8	5.2	2090	1 V38564	Human nuclear prot
38	34.8	5.2	2090	1 V68588	Nucleotide sequenc
39	34.6	5.2	2866	1 X12987	Enterococcus faeca
c 40	34.4	5.1	736	1 T14225	Bettie patched gen
c 41	34.4	5.1	736	1 V21586	Tribolium castaneu
c 42	34.4	5.1	736	1 V64098	Tribolium castaneu
43	34.4	5.1	1291	1 T86248	ACC oxidase GEFE-1
44	34.4	5.1	5100	1 X20271	Borrelia burgdorfe
c 45	34.2	5.1	1660	1 T94661	Chrysanthemum flav

ALIGNMENTS

RESULT 1

V63516

ID V63516 standard; cDNA; 1299 BP.

AC V63516;

DT 28-JAN-1999 (first entry)

DE Consensus sequence of CSI41 ESTsense.

KW EST sequence; CSI41; gastrointestinal tract; cancer; Barret's esophagus;

KW gastric ulcer; gastritis; leiomyoma; polyyps; Crohn's disease;

KW ulcerative colitis; pancreatitis; ss.

OS Synthetic.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 286..957

FT /*tag= a

PN WO9844133-A1.

PD 08-OCT-1998.

PF 31-MAR-1998; U06337.

PR 31-MAR-1997; US-828489.

PA (ABBO) ABBOTT LAB.

PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,

PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,

PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

DR WPI: 98-568280/48.

DR P-PSDB; W80475.

PT New gastrointestinal tract specific polynucleotides, CSI41 - used to develop products for the diagnosis and treatment of e.g. cancers, gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or pancreatitis

PS Claim 1; Page 93; 116pp; English.

CC The present sequence is a consensus sequence of a set of contiguous and partially overlapping EST sequences designated CSI41. The sequences are isolated from a cDNA library made from gastrointestinal tract

CC tumour and normal tissues. The CSI41 gene is useful as a marker for gastrointestinal tract disorders. The methods and products can be used in detecting, diagnosing, staging, monitoring, prognosticating,

CC preventing, or determining the predisposition to diseases and conditions of the gastrointestinal tract, such as gastrointestinal

CC tract cancers, Barret's esophagus, gastric ulcer, gastritis,

CC leiomyoma, polyyps, Crohn's disease, ulcerative colitis, pancreatitis.

CC Sequence 1299 BP; 303 A; 353 C; 383 G; 259 T;

SQ

Query Match

Best Local Similarity 72.6%; Score 485.8; DB 1; Length 1299;

Matches 399; Conservative 152; Mismatches 118; Indels 0; Gaps 0;

QY 1 atgmngntnwsngngntnytnmngnytnngcnytnaththtgcgnathgtacnactgg 60

Db 286 ATCAGAGTGTGACGTGTGCTTCCCTCCCTCATCTTTGCCATAGTCACGACATGG 345

QY 61 atgttyathmgnwsntayatgwsnttywsnatgaaracnathmngnytnccnmngntggytn 120

Db 346 ATGTTTATTTCGAAGCTACATGAGCTTTCAGCATGAAACCATCCGTCTGCCACGCTGGCTG 405

QY 121 gcnwsnccnnaargathcargtgnaaaraatayaartgyggnynathaarccntgy 180

Db 406 GCCTCCGCCACCAGGAGATCCAGGTTAAAGTACAGTGTGGCTCTCATCAAGCCCTGC 465

of which, in non-colon tissue samples, can be used as indication of colon cancer metastasis

Claim 1; Fig 4; 60pp; English.

CC 13 cDNA clones (T45880-92), most of them partial clones, corresponding to human colon specific genes, designated CSG1, CSG2, etc., that are primarily expressed in tissues derived from the colon. CSG7 and CSG10 show reduced expression in colon cancer cells as compared to that in normal cells; the remaining genes are overexpressed in colon cancer. The partial cDNA sequences can be used to isolate full-length clones and genomic clones including the complete gene.

CC CSG nucleic acids can be used to produce CSG polypeptides (see also CC W06545-53) in transformed host cells, as probes to detect disorders of the colon, partic. colon cancer and colon cancer metastasis, and in gene therapy.

Sequence 548 BP. 137 A: 172 G: 107 T:

Query Match	44.0%	Score 294.6;	DB 1;	Length 548;
Best Local Similarity	59.3%;	Pred. No. 2.8e-75;		
Matches 240;	Conservative 93;	Mismatches 72;	Indels 0;	Gaps 0;

QY	265	atqwsnccngttnaaraaayaaytgngnmngcnvntnaayaathbcnyntngtnaaygnacn	324
DB	1	ATGAGTCCTGTTGAAAACAATGTGGCGAGAGCCCTAAACATCGCCCTGGTGAATGAAGACC	60
QY	325	acngngngcngtntngncaraargcnttygayatgtaywsngngngaygtnatgcyyttn	384
DB	61	ACGGGAGCTGTGCTGGGACAGAAGGCATTTGACATGTACTCTGAGATGTTATGCACCTA	120
QY	385	gttnaarttytnaargaratcncngngngcnyntngntnyntngtncnwsntaygayay	444
DB	121	GTGAATTTCTTTAAAGAAATTCGGGGGGTGCATGTGTGTGGCTCTACGACGAT	180
QY	445	ccngngnacnaaratgaaygayarwsnmgnaaryntntywsndaytytngngnwsnwsntay	504
DB	181	CCAGGGACCAAAATGAACGATGAAGACGAGAACTCTTCTGACTGGCGAGTCTCTAC	240
QY	505	gcnaaarcytagngnttymngaygyswtnggnttlyathgngcnnaargaytytmngngn	564
DB	241	GCAAAAACAATCGGGCTTCGGGACAGCTGGGTCTTTCATAGGAGCCAAAGACCTTCAGGGGT	300
QY	565	aarwsnccnttygarcarttytnaaraaywsncngayacnaayaartaygargntgntg	624
DB	301	AAAAGCCCTTTTGAGCAGTTCTTTAAAGAACACGCCACGACACAAACAATACGAGGGATCG	360
QY	625	ccngarytntngaratgargngntgyatgcnccncaarcentty	669
DB	361	CCAGAGCTCTGGATGAGGGGTGATGTTTTCCTCCCGGAGGCATTT	405

```

RESULT      4
V16671
ID          VI6671 standard; cDNA; 548 BP.
AC          V16671;
DT          22-JUN-1998 (first entry)
DE          Polynucleotide sequence of a colon-specific gene.
KW          Colon-specific gene; probe; detection; expression; human;
KW          diagnostic assay; colon cancer; antibody; screening; ss.
OS          Homo sapiens.
FH          Key
FH          Location/Qualifiers
FT          1. .406
FT          /*tag= a
FT          US5733748-A.
PN          31-MAR-1998.
PD          31-MAR-1998.
PF          06-JUN-1995; 469667.
PR          06-JUN-1995; US-469667.
PA          (HUMA-) HUMAN GENOME SCI INC.
PI          Rosen C, Yu G;
PI          WPI; 98-229823/20.
DR          P-PSDB: W46878.
DR          P-PSDB: W46878.
PT          Colon-specific nucleic acids - useful as probes for detecting colon
PT          cancer micrometastases
PS          Claim 15; Fig 4; 51pp; English.

```

CC 16668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes, e.g. in
CC diagnostic assays for detection of the metastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics. 173 G: 107 T:
CC Sequence 548 bp: 137 A: 128 C: 173 G: 107 T:
CC

Query Match	44.0%	Score	294.6	DB	1	Length	548
Best Local Similarity	59.3%	Pred. No.	2.8e-75				
Matches	240	Conservative	93	Mismatches	72	Indels	0
Gaps	0						
Oy	265	atgwsncngnnaaraayaaygtngmgngngnytnaayathgcnyngtngtnaayggnaacn	324				
Db	1	ATGAGTCTGTGAAAAACAATGTGGCGAGAGCCCTAAACATCGCCCTGGTGAATGGGAACC	60				
Oy	325	acngdngcngtnytnngncnaraargcnttygayatgtaywsngngaygvtgnatgcaytn	384				
Db	61	ACGGGAGTGCTGCTGGGACAGAAGGCAATTGCATGTACTCTGGGAGATGTTATGCACCTA	120				
Oy	385	gtinaarttyynaargarathccngngngcngtnytnngtnytnngcwnsntaygayay	444				
Db	121	GTGAAATTCCTTAAGAAATTTCCGGGGGGTGCACTGGTGTGGTGGGCTCTACGACGAT	180				
Oy	445	ccngdgnacnaaratgaaygayarwsmngnaaryntnttywsngaytytnngngnwsntay	504				
Db	181	CCAGGGACCAAAATGAAGGATGAAAGCAGGAAACTTCTCTGACTTGGGAGTTCCTAC	240				
Oy	505	gcnaraarytngnttyngngaywsntggntnttyathgngcnaargayyytnmgngn	564				
Db	241	GCAAAACAACATGGGCTTCCGGGACAGCTGGGTCTTCATAGGAGGCCAAAGACCTCAGGGGT	300				
Oy	565	aarwsncnttygarcatrtytnnaaraaywsncngayacnaayaartaygaagntgg	624				
Db	301	AAAAGCCCTTTGAGCAGTCTTTAAGACACAGCCACGACACAAACAATACGAGGGATGG	360				
Oy	625	ccngarytnytnaratggargntgtgatccnccnnaarcntty	669				
Db	361	CCAGAGTGCTGGAGATGGAGGGCTGATGCCCGAAGCACTTTT	405				

RESULT	5	
V63509		VD3509 standard; cDNA; 263 BP.
ID	AC	V63509;
DT	28-JAN-1999	(first entry)
DE	Clone 1281865	EST corresponding to CS141 sequence.
KW	EST sequence; CS141;	gastrointestinal tract; cancer; Barret's esophagus;
KW	gastric ulcer; gastritis;	leiomyoma; polyyps; Crohn's disease;
KW	ulcerative colitis;	pancreatitis; ss..
OS	Synthetic.	
OS	Homo sapiens.	
PN	W09844133-Al.	
PD	08-OCT-1998.	
PF	31-MAR-1998;	U06337.
PR	31-MAR-1997;	US-828489.
PA	(ABBO)	ABBOTT LAB.
PI	Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,	
PI	Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,	
PI	Kratovichil JD, Roberts-Rapp L, Russell JC, Stroupe SD;	
DR	WPI; 98-568280/48.	
PT	New gastrointestinal tract specific polynucleotides, CS141 - used to	
PT	develop products for the diagnosis and treatment of e.g. cancers,	
PT	gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or	
PT	pancreatitis	
PS	Claim 1; Pages 90-91; 116pp; English.	
CC	V63504-15	represent a set of contiguous and partially overlapping EST
CC	sequences designated CS141. The sequences are isolated from a cDNA	
CC	library made from gastrointestinal tract tumour and normal tissues.	
CC	The CS141 gene is useful as a marker for gastrointestinal tract	
CC	disorders. The methods and products can be used in detecting,	


```
PF 31-MAR-1998; U06337.  
PR 31-MAR-1997; US-828489.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
P1 Gordon J, granados EN, Hayden M, Hodges JC, Klass MR,  
F1 Kratochvil JD, Roberts-Rapp L, Russell SC, Stroupe SD;  
DR WPI : 98-568280/48.  
PT New gastrointestinal tract specific polynucleotides, CS141 - used to  
develop products for the diagnosis and treatment of e.g. cancers,  
gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or  
pancreatitis  
PS Claim 1; Pages 91-92; l16pp; English.  
CS V03504-15 represent a set of contiguous and partially overlapping ES  
sequences designated CS141. The sequences are isolated from a CDNA  
library made from gastrointestinal tract tumour and normal tissues.  
CC The CS141 gene is useful as a marker for gastrointestinal tract  
disorders. The methods and products can be used in detecting, or  
diagnosing, staging, monitoring, prognosticating, preventing, or  
determining the predisposition to diseases and conditions of the  
gastrointestinal tract, such as gastrointestinal tract cancers,  
Barret's esophagus, gastric cancer, gastritis, leiomyoma, polyyps,  
Crohn's disease, ulcerative colitis, panreatitis.  
CC Sequence 187 BP;   35 A;    42 C;   74 G;   33 T;  
SQ
```

```
Query Match          6.0%; Score 40.4; DB 1; Length 187;  
Best Local Similarity 67.3%; Pred.No. 0.018;  
Matches 35; Conservative 9; Mismatches 8; Indels 0; Gaps
```

```
QY      618 rgngtgcgcggaryntnyngatrgtgargnqyvatccncnaarcentty 669  
       ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db      1 GGATGCCAGAGCTGCTGGAGATGGAGCGTGATCCCCGAAGCCAATT 52  
       |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
```

```
RESULT 10  
N70938  
ID ID N70938 standard; DNA; 3830 BP.  
IC IC N70938;  
DT DT 10-APR-1991 (first entry)  
DE DE DNA fragment contig. streptomycin resistance gene and its regulatory  
sequence from Streptomyces griseus 4-1 strain (FERM BP-1198).  
KW KW Antibiotic-resistance; selection marker; ss.  
OS OS Streptomyces griseus.  
FH FH Key Location/Qualifiers  
FT FT cds 1043..1966 /tag= a
```

```
PB P08703302-A.  
PN PN 04-JUN-1987.  
PD PD 20-NOV-1986; J00596.  
PF PF 20-NOV-1985; JP-258622.  
PA (ZAID ) ZH BISEIBUTSU KAGAKU KEN.  
PA (UMEZ.) UMEZAWA H.  
PI Unezawa H, Ikami Y;  
FI WPI: 87-163322/23.  
DR DR DNA contg. streptomycin resistance gene and regulatory sequence -  
used as resistance selection marker, isolated from plasmid pST141  
Streptomyces lividans 4-1  
PS Disclosure; Fig 3; 4lp; Japanese.  
CC The DNA fragment can be inserted into an actinomycetes plasmid vector (e.g. pIJ720) to give a hybrid plasmid which may be used as a streptomycin resistance selection marker.
```

```
SQ Sequence 3830 BP;   576 A;   1367 C;   1348 G;   538 T;
```

```
Query Match          5.8%; Score 39; DB 1; Length 3830;  
Best Local Similarity 27.0%; Pred.No. 0.46;  
Matches 89; Conservative 49; Mismatches 192; Indels 0; Gaps
```

```
QY     195 ygenttaarathtgywsgngcngcnuaayvgtngtngnccnacnactgtgttytgarga 250  
       ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     17 CGCCACCAAAATGC GCGGACCACCGAGATCGC GTGTTC GAAGCGGAGAGA 76  
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
```

```

Db 77 AGAGCGGTTTCATCTTCGCGGTGAAGTCCACGACTCACCCACGAGCTCCGCTCTCCCTCGA 136
Qy 315 naaygnacnagngcngntgntgngcncaraargcnttygayatgtaywsngngaygt 374
Db 137 CGACCCGACGAGCGCGGCGACCGCGTCTCTGGAGACCCATCCGCTCTGGTCCGACCGGGC 196
Qy 375 natgcaytngtnaarttytynaagarathccngngngcngntgntgntgntgncnws 434
Db 197 CATCGGCTTCGGCGACGGACTGTCGCGGAAGAGCGGTGGGACCCCTCAGGGTCTCTTCGA 256
Qy 435 ntaygsgyccngcngnacnaaratgaaygayarwsnmgnaaryntntysngaytngg 494
Db 257 CTGCGCGGTTCGCGAGTTCGACGATCGAGATCGGAGGCGGCGGCGCGCCCGCTCGA 316
Qy 495 nwsnwtaygcnacnarytngnttymg 524
Db 317 CCCACCGAGGCGGCGAAGCTGCGGAGCGG 346

RESULT 11
V74910
ID V74910 standard; DNA; 704 BP.
AC V74910; 1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #599.
KW Computer readable medium; vaccine; S.aureus infection; Immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 361..420
FT /*tag= a
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
EP-786519-A2.
30-JUL-1997.
07-JAN-1997; 100117.
05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI: 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1531-1532; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 704 BP; 232 A; 98 C; 101 G; 213 T;

```

Query Match 5.7%; Score 37.8; DB 1; Length 704;
 Best Local Similarity 28.8%; Pred. No. 0.28;

```

Matches 78; Conservative 32; Mismatches 161; Indels 0; Gaps 0;
Qy 17 tnytnmgyntngcnytnathttgcnathgtgnacnctggtgattgttyatghmgnwnt 76
Db 354 TTTTCATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 413
Qy 77 ayatgswnttysnatgaaracnathmgnytncnmgntggytngcngnwsncnacnaarg 136
Db 414 NNNNNNNNTATTAGTATTAGGTTATGTTATTTATATTGCTAGCAAGACTTCCACCAATT 473
Qy 137 arathcargtnaaraatayaartgyggnvtnatharccntgycngcnaaytayttyg 196
Db 474 AATCAAGCCCAACAATAAACTGTTGTATATAATACCTTATTTCCAAATTTTCAG 533
Qy 197 cnttyaarathtgysngngcngcnaaytngntgngcncnacnctgtytgygargaym 256
Db 534 AATAATAACATTGTTAAATGAGCTGTTGCACAGTGCATAATAGTAATAAATCGATAATA 593
Qy 257 gnatgathatgwsncngtnaaraayaaygt 287
Db 594 GCATTAATAGATAAAACGAGATAAATCATCT 624

RESULT 12
V69614/c
ID V69614 standard; DNA; 1341 BP.
AC V69614;
DE 28-JAN-1999 (first entry)
KW Human secreted protein gene 4 clone HBND556.
KW Secreted protein; gene therapy; protein therapy; diagnosis; treatment;
KW central nervous system; CNS; immune system; cancer; trauma; liver;
KW reproductive disorder; congenital malformation; degenerative disease;
KW inflammatory disease; neoplasia; metabolic disorder; testis; placenta;
KW brain; T cell; spleen; lung; heart; rhabdomyosarcoma; endocrine system;
KW endocrinopathy; endocrine polylglandular syndrome; endocrinoma; sepsis;
KW endocrine ophthalmopathy; osteoclastoma; bacterial infection; bone; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 530..745
FT /*tag= a
FT /*product= "secreted protein"
FT sig_peptide 530..604
FT /*tag= b
FT mat_peptide 605..742
FT /*tag= c
FN WO9845712-A2.
PD 15-OCT-1998.
PE 07-APR-1998; U06801.
PR 30-MAY-1997; US-046184.
PR 08-APR-1997; US-042726.
PR 08-APR-1997; US-042727.
PR 08-APR-1997; US-042728.
PR 08-APR-1997; US-042754.
PR 08-APR-1997; US-042825.
PR 30-MAY-1997; US-048068.
PR 30-MAY-1997; US-048070.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Feng P, Ni J, Rosen CA, Ruben SM, Yu G;
PI WPI: 98-594456/50.
DR P-FSD; W83934.
DR New isolated human genes and secreted polypeptide(s) they encode -
PT useful for the diagnosis and treatment of e.g. cancers, CNS
PT disorders, immune system disorders, inflammatory disease and
PT bacterial infections
PS Claim 4; Page 104; 142pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 4 from
CC the human cDNA clone HBND56 (deposited as clone ATCC 97955 and ATCC
CC 209074) which encodes a secreted human protein. This gene maps to
CC chromosome 4 and is abundantly expressed in parathyroid tumour. It is
CC useful for diagnosis and treatment of disorders of the nervous system and
CC the protein products are useful in modulating calcium metabolism.
CC The invention relates to 20 novel genes and their fragments (V69611 to
CC V69630) and corresponding secreted proteins (W83931 to W83950) which are

```


CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein of gene therapy. Also pathological conditions can be diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the polynucleotides. Specific
CC uses are based on which tissues they are most highly expressed in, and
CC include developing products for the diagnosis or treatment of central
CC nervous system (CNS) and immune system diseases, reproductive disorders,
CC cancers, congenital malformations, degenerative diseases, trauma,
CC inflammatory disease, neoplasia, metabolic disorders, diseases in testes,
CC placenta, liver, brain and activated T cells, spleen diseases, lung
CC diseases, heart diseases, rhabdomyosarcoma and disorders of the endocrine
CC system or other endocrinopathies, e.g. endocrine polyglandular syndrome,
CC endocrinoma, and endocrine ophthalmopathy, osteoclastoma and other bone
CC remodelling disorders, bacterial infections and sepsis. The polypeptides
CC are also useful for identifying their binding partners. The polypeptides
SQ Sequence 1341 BP; 376 A; 278 C; 313 G; 373 T;

Query Match 5.5%; Score 37; DB 1; Length 1341;
Best Local Similarity 28.3%; Pred. No. 0.77;
Matches 91; Conservative 60; Mismatches 167; Indels 4; Gaps 2;

QY 282 yaaygtngngmgngnytnaayathgncyngntngnaaygngnacngngcngntngng 341
DB 553 CAACGTAGGAAGAGAGACAGAAACATCTCTCTCCACCTTATACCTGCCACTGTGTT--- 496
QY 342 ncaaraagcnttygatytaywsngngngaygnatgcaaytngntnaarttytnaarga 401
DB 496 TTAAGAGGTATATGGTATTAAAGAAAGTTGGCTGTGTTGCTTTTAAATTTTAAATTT 437
QY 402 rathcngngngcngnytnngntngncnwsntaygayccngnacnaaratgaa 461
DB 436 ATGAACTAGCGCTGCCACTCTCTCTGCTTCTTACTTTGAGTGTCATCTAGCTGCC 377
QY 462 Ygaygarwsnmgnaarytnnt-ywsngaytngngnwsntaygnaacarytngngnt 520
DB 376 TGTGGTATCATATTAGGGATCCCATCAATGATGGATAAGCTATTCCCAACTCTTCATTA 317
QY 521 tynngaywsntgggtnttyathgngcnaargaytngngngnacnwsncnttygarc 580
DB 316 ATCAATTCGTTTGTGTTGCTTATCTGAGCGGCTTCTTGAGAGCGGCGACACCGAG 257
QY 581 arttytnaaraaywsncnga 602
DB 256 AACTCCAGCAGCCCGGATCGA 235

RESULT 13
T08144
ID T08144 standard; DNA; 1633 BP.
AC T08144;
DT 12-MAY-1996 (first entry)
DE Xylanase xynB gene (partial sequence).
KW Xylanase B; xynB gene; thermostable enzyme; paper; pulp; lignin;
KW delignification; xylan; bleaching; ds.
OS Extremophile isolate TG456 (CBS 213.94).
PN W09534662-AL.
PF 21-DEC-1995.
PP 14-JUN-1995; E02299.
PR 14-JUN-1994; EP-201699.
PA (KONN) GIST-BROCADES BV.
PI Bergquist PL, Daniel RM, Farrel RL, Forster S, Gronberg V;
PI Herweller MA, Iverson S, Jones BE, Moody D, Morgan HW;
PI Quax WJ, Williams DP;
DR WPI; 96-049690/05.
DR P-PSDB; R87014.
PT Novel xylanase(s) having activity at 80 deg C. or higher - obtained
PT from anaerobic thermophilic bacteria and used in paper and pulp
PT production processes
PS Example 8; Page 47-49; 77pp; English.
CC A xynB gene partial sequence (T08144) codes for a thermostable F-type
CC xylanase (R87014) useful in the pulp and paper industries. The
CC xynB gene, which codes for a multi-domain enzyme that includes a

CC single xylanase B domain, was obtd. by genomic walking PCR of DNA from
CC extreme thermophilic strain TC456 (CBS 213.94), isolated from a New
CC Zealand hot spring. The gene can be inserted into a vector and used
CC for the prodn. of recombinant xylanase B in microbial host cells, esp.
CC Escherichia coli.
SQ Sequence 1633 BP; 564 A; 291 C; 347 G; 431 T;

Query Match 5.5%; Score 36.8; DB 1; Length 1633;
Best Local Similarity 27.9%; Pred. No. 1;
Matches 57; Conservative 41; Mismatches 106; Indels 0; Gaps 0;

QY 449 gnaacnaaratgaaygaygarwsnmgnaarytnntysngaytngngnwsntaygna 508
DB 409 GCACAGACCAATATATTTGGATTAAGAGGTCATACACTGGTTGGCATCAACAACCTCCA 468
QY 509 arcarytngngntymngngaywsntgggtnttyathgngcnaargaytngngngnaarw 568
DB 469 GATTGGTTTTTCCAGCATCTCTGACGGTTCCGCACTTGATCCAAACAATCTTGAAGACAAG 528
QY 569 snconttygarcarttytnaaraaywsncngayacnaayaartaaygarngntggcng 628
DB 529 CAGCTTTTTCGAAAATAGGTTAAACACACACATTCAGACACTTGTGTGGAAGATATGCACAG 588
QY 629 arytntngaratgargngtgya 652
DB 589 AAAGTTTATGATGGGATGTTGTA 612

RESULT 14

V31480
ID V31480 standard; DNA; 987 BP.
AC V31480;
DT 08-SEP-1998 (first entry)
DE Human interleukin-12 (IL-12) p40 subunit encoding DNA.
KW Interleukin-12; IL-12; p40 subunit; p35 subunit; human; asthma;
KW cancer; therapeutic gene delivery; vaccine; allergy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Key 1. 987
FT CDS
FT /product= "human IL-12 p40 subunit"
FT /note= "the protein is encoded by degenerate codons
FT which can code for each of the amino acid"

W09817814-A2.
PD 30-APR-1998.
PP 10-OCT-1997; U18832.
PR 18-OCT-1996; US-028687.
PA (GENE-) GENEMEDICINE INC.
PI Deshpande D, Freimark B, Nordstrom J;
DR WPI; 98-261507/23.
DR P-PSDB; W57482.
PT Plasmid for co-ordinated expression of recombinant eukaryotic genes
PT - useful for, e.g. therapeutic gene delivery, especially for
PT treating asthma and cancer
PS Disclosure; Fig 4A-D; 104pp; English.
CC This DNA encodes a human interleukin-12 (IL-12) p40 subunit. This is used
CC in the construction of plasmid for expressing recombinant eukaryotic
CC genes. The plasmid comprises a first transcription unit comprising
CC control sequences linked, in order, to a 5'-untranslated region (UTR),
CC synthetic intron, a coding sequence and synthetic 3'-untranslated region
CC /polyA sequence (UTR/A), and a second transcription unit of similar
CC construction. The first coding sequence and second coding sequence
CC comprise a sequence coding for p40 subunit of human IL-12 and a sequence
CC coding for p35 subunit of human IL-12. The plasmids are used for
CC therapeutic gene delivery, especially for treating asthma and cancer.
CC They are also useful as adjuvants for vaccines. IL-12 is known to
CC stimulate natural killer cells, to promote maturation of cytotoxic T
CC cells and to induce cytokine production by these cells. They stimulate
CC cell-mediated immune responses in cases of microbial or viral infection,
CC some cancers, allergies and asthma. Co-ordinated expression of two
CC subunits from a single plasmid is now possible, avoiding the difficulties
CC associated with transfection with two separate plasmids.

```
SQ Sequence 987 BP; 181 A; 88 C; 165 G; 125 T;

Query Match
Best Local Similarity 5.5%; Score 36.6; DB 1; Length 987;
Matches 60; Conservative 154; Mismatches 387; Indels 0; Gaps 0;

QY 1 atgmgngtwsngngntnmgntnnytnngcnytnathtttgcgnathgtnacnctgg 60
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 AYACNTGICAYAAAGGNGGARGTNYTNWNCAYWSNYTNNTNTNCAYAAARG 322
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 atgttyatmgwntayatgnttysnatgaaracnathmgntnccmngtggyn 120
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 ARGAYGGNATHTGGWSNACNGAYATHYTNAAARGAYCARAARGARCNAARAAYARACNT 382
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 gcnwsnccnacnaargatgathcarytgnaaartaartayartgygnytnaathaaacctgy 180
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 TYTNTMGNTGYGARGCNAARAAYATYWSNGNGMTTACNTGYTGGGGYTNACNACNA 442
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 cngcnaaytaytytgcnttyaarathgtgywngngcngcnaaygtngtngngcncn 240
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 THWSNACNGAYVTNACNTTYWSNGTNAARWSNWSNMNGNWSNGAYCCNCARGNG 502
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 atgttytgaygmgngatgathatgwsnccngttnaaraayaaygtngmgngnytn 300
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 TNACNTGYGGNGCNCNACNTYNWSNGCNGARMGNGTNGMGNGGAYAYAAARGARTAYG 562
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 aavathgcnnyngtnaayggnacnccngcngcngntnngncaraargcnttygayatg 360
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 ARTAYWSNCTNGARTGYCARGARGAYWSNGCNGTGYCCNGCNGCARGARWSNTYTN 622
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 361 taywngngngaygtngatgavtngttnaarttyytnaargatgcccngngngcnytn 420
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 THGARGTNGATGNGAYGCGNGTNCAYAAARTYTNAAARTAYGARAAYTAYACNWSN 682
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 421 gntnyngtngcwnstaygagaycngcngnacnaaratgagaygarwsmngnaarytn 480
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 TYATHMGNGAYATHAARCCNGAYCCNCCNAARAAYTYTNCARYTNAARCN 742
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 481 ttwswngayvtyngngwswntaygcnaarcarytngnttymngngaywsntggntty 540
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 743 AYSNMGNCARGTNGARGTNSNTGGGARTAYCCNGAYACNTGGWSNACNCCNCAIWSNT 802
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 541 athgngcnaargayvtytmngngngnaarwsnccnttygarcarttyytnaaraaywsnccn 600
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 AVTTYWSNTNACNTTYTGYGTNCARGTNCARGGNAARWSNAARMGNGARAARAARGAYM 862
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 601 g 601
Db 863 G 863
```

RESULT 15

```
T30309/c
ID T30309 standard; cDNA; 3934 BP.
AC T30309;
DT 20-AUG-1996 (first entry)
DE Human SOX-9 cDNA.
KW Sox-9; bone regeneration; cartilage regeneration;
KW campomelic dysplasia; gene therapy; sex reversal; ss.
OS Homo sapiens.
FH key
FT Location/Qualifiers
   cds
   FT 360..1889
   FT /*tag= a
   FT 3896..3901
   FT /*tag= b
   FT
   PN W09617057-A1.
   PD 06-JUN-1996.
   PE 29-NOV-1995; AU0799.
   PR 29-NOV-1994; AU-009714.
   PR 05-DEC-1994; AU-009835.
   PA (UYCA-) UNIV CAMBRIDGE.
   PA (UYQU-) UNIV QUEENSLAND.
```

```
PI Goodfellow PN, Koopman PA;
DR WPI; 96-27777/28.
DR P-PSDB; R97646.
PT New isolated SOX-9 genes - used to develop prods. for the promotion
PT or suppression of bone or cartilage differentiation or growth
PS Claim 2; Fig 8a; 64pp; English.
CC A composite cDNA sequence (T30309) codes for the human
CC transcription factor, SOX-9 (R97646). It was deduced from cDNA
CC clones obtd. from a testis cDNA library using a SOXA HMG (high
CC mobility gp.) probe. The gene appears to play a role in skeletal
CC development. It has been mapped to chromosome 17 within a region
CC which contains CMPD1, the locus for campomelic dysplasia (CD).
CC Alterations in SOX-9 can cause both CD and male to female sex reversal.
CC The cDNA can be used in the prepn. of recombinant SOX-9 polypeptide.
CC DNA or recombinant protein may be injected into joint tissue for
CC treatment of cartilage or bone damage. SOX-9 can also be used as a
CC tool for studying bone formation and sex determination.
SQ Sequence 3934 BP; 966 A; 1100 C; 901 G; 967 T;

Query Match 5.5%; Score 36.6; DB 1; Length 3934;
Best Local Similarity 27.4%; Pred. No. 2.3;
Matches 61; Conservative 43; Mismatches 119; Indels 0; Gaps 0;

QY 424 ytngtngcwnstaygagaycngcngnacnaaratgaaygarwsmngnaarytn 483
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2840 TTGCTTGCCCTTTTACCTCTTTCTTAAGACTGCAGTGAACAAGCAAGGCAGGAGAAATG 2781
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 484 wsgayvtyngngwswntaygcnaarcarytngnttymngngaywsntggnttyath 543
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2780 CACTAAAAGAGTGCAAATGTTTCCAGCAGCACCGTTTAAAGGCTCAAGGTGTTTCTC 2721
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 544 gngcnaargayvtyngngngnaarwsnccnttygarcarttyytnaaraaywsnccngay 603
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2720 TTCTTTAAAAAATATATATATATCAATAAGCTTACCATAATGCTTCTTAGACTATTGTTTAC 2661
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 acnaayaartaygargntggcngarytntngaratgarg 646
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2660 TCCCATATTTGCAAAATAAGGCCTTTTATTTTAAAAAGGCAG 2618
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: May 31, 2000, 19:04:01
Job time: 5944 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 18:37:16 ; Search time 1292.59 Seconds
(without alignments)
-503.483 Million cell updates/sec

Title: US-09-186-342-8
Perfect score: 669
Sequence: 1 atmgngntwsngngntnytt.....gyatgccnccnaarcntty 669

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_em.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*
51: gb_pr5.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	294.6	44.0	548	5	I95745	I95745 Sequence 6
2	260.6	39.0	1606	9	HS219MRNA	X87193 H.sapiens m
3	248.2	37.1	2475	9	D87120	D87120 Homo sapien
c 4	189.8	28.4	465	12	MMU72677	U72677 Mus musculu
c 5	90.8	13.6	178137	55	AC023346	AC023346 Homo sapi
c 6	76.2	11.4	52173	9	HSG6PDGEN	X55448 H.sapiens G
c 7	76.2	11.4	219447	40	HUMFLNG6PD	L44140 Homo sapien
c 8	61.4	9.2	132319	40	AC006364	AC006364 Homo sapi
c 9	52	7.8	37777	34	CELM70	AF047661 Caenorhab
c 10	52	7.8	151727	41	AC006894	AC006894 Caenorhab
c 11	48.6	7.3	7218	5	I66494	I66494 Sequence 14
c 12	48.2	7.2	1302	51	AF090946	AF090946 Homo sapi
c 13	48	7.2	6031	1	CBGIDPAB	Y10436 C.burnetii
c 14	44	6.6	222877	32	AL133377	AL133377 Homo sapi
c 15	43.2	6.5	249287	41	AC006745	AC006745 Caenorhab
c 16	42.6	6.4	69214	43	AC018572	AC018572 Homo sapi
c 17	41.8	6.2	61771	54	AC023101	AC023101 Homo sapi
c 18	41.6	6.2	67994	40	AC006981	AC006981 Homo sapi
c 19	41.6	6.2	74371	11	AC005369	AC005369 Homo sapi
c 20	41.6	6.2	175986	42	AC010126	AC010126 Homo sapi
c 21	41.2	6.2	5527	34	TETTRAG	M59429 T. thermoph
c 22	41.2	6.2	84630	45	AC008933	AC008933 Homo sapi
c 23	41.2	6.2	136862	40	AC005007	AC005007 Homo sapi
c 24	41.2	6.2	183516	44	AC018368	AC018368 Homo sapi
c 25	41	6.1	1409	8	AF016100	AF016100 Arabidops
c 26	41	6.1	12203	1	SGSTR	Y00459 Streptomyce
c 27	41	6.1	19292	34	CEY76A2A	AL032865 Caenorhab
c 28	41	6.1	269082	32	CEY76A2	Z92866 Caenorhabdi
c 29	40.6	6.1	4161	7	ATHRD21A	D13043 A.thaliana
c 30	40.6	6.1	34598	35	CELT09D3	U64835 Caenorhabdi
c 31	40.6	6.1	41229	34	CEC15H11	Z81035 Caenorhabdi
c 32	40.6	6.1	64168	52	AC008607	AC008607 Homo sapi
c 33	40.6	6.1	79655	44	AC008724	AC008724 Homo sapi
c 34	40.6	6.1	133599	44	AC011339	AC011339 Homo sapi
c 35	40.6	6.1	181923	52	AC011741	AC011741 Homo sapi
c 36	40.6	6.1	226574	32	CEY49A3	AL023793 Caenorhab
c 37	40.2	6.0	2113	34	AB017107	AB017107 Caenorhab
c 38	39.8	5.9	153507	43	AC015823	AC015823 Homo sapi
c 39	39.8	5.9	165766	58	AC018361	AC018361 Homo sapi
c 40	39.8	5.9	169889	44	AC016356	AC016356 Homo sapi
c 41	39.8	5.9	171967	44	AC011678	AC011678 Homo sapi
c 42	39.8	5.9	191660	33	AL138718	AL138718 Homo sapi
c 43	39.6	5.9	99670	40	AC007200	AC007200 Homo sapi
c 44	39.6	5.9	99670	41	AC008677	AC008677 Homo sapi
c 45	39.6	5.9	203662	42	AC011235	AC011235 Homo sapi

ALIGNMENTS

RESULT 1	195745	I95745	548 bp	DNA	PAT	01-DEC-1998
LOCUS	195745	Sequence 6 from patent US 5733748.				
DEFINITION	195745	Sequence 6 from patent US 5733748.				
ACCESSION	195745	GI:3940215				
VERSION	195745.1	GI:3940215				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 548)					
AUTHORS	Yu, G. and Rosen, C.					
TITLE	Colon specific genes and proteins					
JOURNAL	Patent: US 5733748-A 6 31-MAR-1998;					
FEATURES	Location/Qualifiers					
source	1. 548					
BASE COUNT	137 a	128 c	173 g	107 t	3 others	
ORIGIN	/organism="unknown"					
	44.0%; Score 294.6; DB 5; Length 548;					
	Best Local Similarity 59.3%; Pred. No. 1.4e-61;					
	Matches 240; Conservative 93; Mismatches 72; Indels 0; Gaps 0;					
QY	265 atgwsncnctnaaraayaagvtngnngnngnytnaavathgcnytgntnaaygnaacn	324				
Db	1 ATGAGTCTGTGAAAACAATGTGGCGAGAGGCTTAACATGCCCTGGTGAATGGAACC	60				
QY	325 acngngcngtntngnncaraargcnttygayatgtaywsngngaytgntatgcytyn	384				
Db	61 ACGGAGCTGTGCTGGGACAGAGGCAATTTGACATCTACTCTGGAGATGTTATGCACCTA	120				
QY	385 gtnaarttytnaargarathcncngngngcnytgntngtngtngcwnstnaygayay	444				
Db	121 GTGAAATTCCTTAAGAAATTCGCGGGGGTGCACCTGGTCTGTGGCCCTACGACCAT	180				
QY	445 cngngnacnaaratgaaygaygarwsnmgnnaarytnttywsngaytyngnwnstnay	504				
Db	181 CCAGGGACCAAAATGACGATGAAGCAGGAACACTCTCTGACTTGGGGAGTTCCTAC	240				
QY	505 gcnarcarcnytgntymngngaywsntgggtnttyathgngncnaargaytynmngnn	564				
Db	241 GCAAAACAACCTGGGCTCCCGGACAGCTGGGTCTTCATAGGACCAAAACCTCAGGGGT	300				
QY	565 aarwsncnttygarcttytnaaraaywsncngayacnaayaartayargntgg	624				
Db	301 AAAGCCCTTTGAGCAGTTCCTTAAAGACAGCCGACCAACAAATACAGGGGATGG	360				
QY	625 ccngarytntngaratggargnttyatgcncncnaarccntty	669				
Db	361 CCAGAGCTGTGGAGATGGAGGCTGCATGCCCCGAGGCCATTT	405				
RESULT 2	HS219MRNA	1606 bp	mRNA	PRI	02-JUN-1995	
LOCUS	HS219MRNA	Sequence 6 from patent US 5733748.				
DEFINITION	HS219MRNA	Sequence 6 from patent US 5733748.				
ACCESSION	HS219MRNA	GI:854081				
VERSION	HS219MRNA.1	GI:854081				
KEYWORDS	2.19 gene.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 1606)					
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;					
TITLE	Primates; Catarrhini; Hominidae; Homo.					
JOURNAL	1 (bases 1 to 1606)					
MEDLINE	Bione,S., Tamani,F., Maestrini,E., Tribioli,C., Poustka,A.,					
	Torri,G., Rivella,S. and Toniolo,D.					
	Transcriptional organization of a 450-kb region of the human X					
	chromosome in Xq28					
	Proc. Natl. Acad. Sci. U.S.A. 90 (23), 10977-10981 (1993)					
	94068527					

[illegible]


```

exon      17041..17117
/gene="G6PD"
/number=11
variation  /note="near the end of intron 11"
           /citation=[1]
           /replace="c"
exon      17223..17315
/gene="G6PD"
/number=12
exon      17413..18112
/gene="G6PD"
/number=13
variation  17860
/gene="G6PD"
/note="exon13, 3'UTR"
/citation=[1]
/replace="g"
18094..18099
/gene="G6PD"
18112
/gene="G6PD"
/citation=[2]
evidence-experimental
19058
/note="downstream G6PD intergenic region"
/citation=[1]
/replace="t"
33000..33289
/gene="2-19 gene"
/citation=[1]
/label=exon1
/number=1
prim_transcript 33000..42862
/gene="2-19"
/citation=[1]
evidence-experimental
33000..33289
/gene="2-19 gene"
33000..42862
/gene="2-19"
/citation=[1]
/codon_start=1
/label=protein2-19
evidence-experimental
/product="2-19 protein"
/protein_id="CAA39090.1"
/db_xref="GI:452270"
/db_xref="SWISS-PROT:P98173"
/translation="MRLAGPLRIVLVSVGVTVIIVSVILLGGPGSGFPRIQQLFTSP
ESSVAAPARKYKGLPQPCPEEHLAFRVDSGAANVIGPKICLEDKMLMSSVKONVG
RGNIALVNGVSGELTEARAFDMWAGVDLLKFIPLHEGTLVFFVASYDDPATKME
ETRLKLSGLSRKAKELAFRDSWVFVGAQVQNKSPFEQHVKNKSNKYEGCPALE
MECCIPRRSTAS"
36113..36226
/gene="2-19"
/citation=[1]
/label=exon2
/number=2
37166..37189
/gene="2-19"
/citation=[1]
/label=exon3
/number=3
40439..40562
/gene="2-19"

```

```

/citation=[1]
/label=exon4
/number=4
40689..40747
/gene="2-19"
/citation=[1]
/label=exon5
/number=5
41175..41225
/gene="2-19"
/citation=[1]
/label=exon6
/number=6
41546..41630
/gene="2-19"
/citation=[1]

Query Match      11.4%; Score 76.2; DB 9; Length 52173;
Best Local Similarity 45.4%; Pred. No. 8.1e-08;
Matches 64; Conservative 37; Mismatches 40; Indels 0; Gaps 0;

QY  444 yccngnacnaaratgaaygagrvsmngnaarytnttywsngayyytngngwnsnta 503
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  41695 CCATGCTGCAGGATGATGAAGAGAGACCAGAAAGCTCTTCAGTGAGCTGGCAGCAGGAA 41754

QY  504 ygcnaarcarytngnttymngaywsntgggtnttthyathgngcnaargayyytmngng 563
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  41755 CGCAAGAGAGCTGGCCTTCGGGACAGAGCTGGTGTGTTCGGGGCCAGGGTGTGCAGAA 41814

QY  564 naarwsnccnttygarcartt 584
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  41815 CAAAGAGCCCTTTGAGCAGGT 41835

```

```

RESULT 7
HUMFLNG6PD/c HUMFLNG6PD 219447 bp DNA PRI 17-MAY-1999
LOCUS Homo sapiens chromosome X region from filamin (FLN) gene to
DEFINITION glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds's.
ACCESSION L44140
VERSION L44140.1 GI:1203968
KEYWORDS 1A gene; 2-19 gene; ABP-280 gene; DNLI1L gene; DNase I-like protein
gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI
gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene;
XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin;
emerly-drefvuss syndrome; filamin; glucose-6-phosphate
dehydrogenase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219447)
AUTHORS Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N.,
Zuo,L., Helner,C., Burrough,F.W., Ripetto,M., Schlessinger,D. and
D'Urso,M.
TITLE Long-range sequence analysis in Xq28: thirteen known and six
candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
G6PD loci
JOURNAL Hum. Mol. Genet. 5 (5), 659-668 (1996)
MEDLINE 96311563
COMMENT Submitted by:
Ellison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln
Center Drive,
Foster City, CA 9404 USA
and
David Schlessinger,
Department of Molecular Microbiology and Center for Genetics in
Medicine
Washington University School of Medicine,
St. Louis MO 63110 USA
e-mail: ellison@genseq.apldbio.com and david@genetics.wustl.edu

```


Db	23250	CTTGAAGAAATGGCTTGACGGCGTGCCGACCGGCTACATTTAGCCGTTGTGAGCTTCGA	23191
Qy	441	Ygacycngnagacnaaratgaaygarwsmgnaarynttysngayyyngngnshws	500
Db	23190	CGAGGCCTCTAATCAGCTGTGCAGATATGGCTCGGCGCATCTTTACGAGATGGGCTCCAG	23131
Qy	501	ntaygcnacnaaryngnttymngaywsgtsgnttntatthgngngcnaargayyytmng	560
Db	23130	TATGATCGATCGGCTCAAAATTCGAGCTTCGTGTTATTCGTCGACAGAAGGCATCGG	23071
Qy	561	ngnnaarwscncttygarca	581
Db	23070	CGCCTACACACCGTTGAAGA	23050
RESULT 10			
AC006894/c			
LOCUS			
DEFINITION	AC006894 151727 bp DNA HTG 26-FEB-1999		
ACCESSION	Caenorhabditis elegans clone Y71G10, *** SEQUENCING IN PROGRESS		
VERSION	***, 19 unordered pieces.		
KEYWORDS	AC006894.2 GI:4309905		
SOURCE	HTG; HTGS PHASE1.		
ORGANISM	Caenorhabditis elegans.		
REFERENCE	Caenorhabditis elegans.		
AUTHORS	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
TITLE	1 (bases 1 to 151727)		
JOURNAL	Waterston, R.H.		
REFERENCE	The sequence of Caenorhabditis elegans clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 151727)		
JOURNAL	Waterston, R.H.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (24-FEB-1999) Genome Sequencing Center, Washington		
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
JOURNAL	MO 63108, USA		
COMMENT	On Mar 1, 1999 this sequence version replaced gi:4263454. * NOTE: This is a 'working draft' sequence. It currently * consists of 19 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.		
	1 4066: contig of 4066 bp in length		
	* 4076: gap of unknown length		
	* 4077 7037: contig of 2961 bp in length		
	* 7047: gap of unknown length		
	* 7048 41941: contig of 34894 bp in length		
	* 41942 41951: gap of unknown length		
	* 41952 48127: contig of 6176 bp in length		
	* 48128 48137: gap of unknown length		
	* 48138 50427: contig of 2290 bp in length		
	* 50428 50437: gap of unknown length		
	* 50438 57580: contig of 7142 bp in length		
	* 57580 57589: gap of unknown length		
	* 57590 60562: contig of 2973 bp in length		
	* 60563 60572: gap of unknown length		
	* 60573 67630: contig of 7058 bp in length		
	* 67631 67640: gap of unknown length		
	* 67641 69902: contig of 2262 bp in length		
	* 69902 69912: gap of unknown length		
	* 69913 83157: contig of 13245 bp in length		
	* 83158 83167: gap of unknown length		
	* 83168 92295: contig of 9128 bp in length		
	* 92296 92305: gap of unknown length		
	* 92306 104058: contig of 11753 bp in length		
	* 104059 104068: gap of unknown length		
	* 104069 109160: contig of 5092 bp in length		
	* 109161 109170: gap of unknown length		
	* 109171 114904: contig of 5734 bp in length		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 31, 2000, 17:04:21 ; Search time 1292.59 Seconds
(without alignments)
-918.913 Million cell updates/sec
Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgccgggaggtgcatct.....aaatttttttgcgtga 1221

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 882769 seqs, -486395729 residues
Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_p11.*
8: gb_p12.*
9: gb_p13.*
10: gb_p14.*
11: gb_p15.*
12: gb_p16.*
13: gb_p17.*
14: gb_p18.*
15: gb_p19.*
16: gb_p20.*
17: gb_p21.*
18: gb_p22.*
19: gb_p23.*
20: gb_p24.*
21: gb_p25.*
22: gb_p26.*
23: gb_p27.*
24: gb_p28.*
25: gb_p29.*
26: gb_p30.*
27: gb_p31.*
28: gb_p32.*
29: gb_p33.*
30: gb_p34.*
31: gb_p35.*
32: gb_p36.*
33: gb_p37.*
34: gb_p38.*
35: gb_p39.*
36: gb_p40.*
37: gb_p41.*
38: gb_p42.*
39: gb_p43.*
40: gb_p44.*
41: gb_p45.*
42: gb_p46.*
43: gb_p47.*
44: gb_p48.*

Result No.	Score	Query Match %	Length	DB ID	Description
1	515.6	42.2	548	5	I95745 Sequence 6
C 2	404.2	33.1	178137	55	AC023346 Homo sapi
3	247.8	20.3	1606	9	X87193 H.sapiens m
C 4	229.4	18.8	465	12	U72677 Mus musculu
5	216.4	17.7	2475	9	D87120 Homo sapien
6	70	5.7	52173	9	X55448 H.sapiens G
C 7	56.6	4.6	132319	40	L44140 Homo sapien
8	46.6	3.8	7218	5	AC006364 Homo sapi
C 9	43.4	3.6	35052	57	I66494 Sequence 14
C 10	43	3.5	1302	51	AC016137 Homo sapi
11	43	3.5	1302	51	AF090946 Homo sapi
C 12	43	3.5	173823	33	HSJ963E22 Homo sapi
C 13	41.4	3.4	123645	33	HSJ697K14 Homo sapi
C 14	41.4	3.4	185436	33	HSJ663D7 Homo sapi
15	39.8	3.3	40969	34	LMFL4768 Homo sapi
16	39.6	3.2	112626	41	AC008616 Homo sapi
17	39.2	3.2	2440	5	A20492 cDNA insert
C 18	39.2	3.2	2440	5	A20493 cDNA insert
19	39.2	3.2	2458	5	A20495 vector pT7/
C 20	39.2	3.2	2458	5	A20496 vector pT7/
21	39.2	3.2	3086	5	A20498 RFX gene. 7
22	39.2	3.2	3086	9	X58964 H.sapiens g
23	38.4	3.1	178896	40	AC008041 Homo sapi
C 24	38.4	3.1	283438	57	AC012465 Homo sapi
C 25	38.2	3.1	173561	32	AP001141 Homo sapi
26	38	3.1	1351	12	AF022079 Mus muscu
C 27	38	3.1	3233	12	AF027181 Rattus no
28	38	3.1	4505	12	AF016252 Rattus no
C 29	38	3.1	6803	11	AF016252 Rattus no
30	37.8	3.1	2976	2	AF226283 Xanthomon
31	37.4	3.1	55590	56	AC011531 Homo sapi
32	37.2	3.0	2454	10	HSSGI Y00084 Human mRNA
C 33	37.2	3.0	120940	55	AC018939 Homo sapi
C 34	37.2	3.0	139352	10	HS967N21 Human DNA
C 35	37	3.0	2282	9	HUMSCLAA M31731 Human chron
36	37	3.0	5200	9	HUMSCLAA M31731 Human chron
37	37	3.0	5247	10	HSU29175 Human trans
38	37	3.0	23550	1	SCE39 AL049573 Streptomy
39	36.8	3.0	160218	45	AC010335 Homo sapi
40	36.4	3.0	74079	45	AC021480 Mus muscu
41	36.4	3.0	163253	43	AC011978 Homo sapi
C 42	36	2.9	59438	42	AC015710 Homo sapi
C 43	36	2.9	216347	33	AL138885 Homo sapi
44	35.8	2.9	380	13	G54682 AffySTS0000
45	35.8	2.9	380	13	G50062 SHGC-130892

ALIGNMENTS

```
RESULT 1
195745
LOCUS 195745 548 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 6 from patent US 5733748.
ACCESSION 195745
VERSION 195745.1 GI:3940215
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 548)
AUTHORS Yu,G. and Rosen,C.
TITLE Colon specific genes and proteins
JOURNAL Patent: US 5733748-A 6 31-MAR-1998;
FEATURES
Location/Qualifiers
1..548
source
/organism="unknown"
BASE COUNT 137 a 128 c 173 g 107 t 3 others
ORIGIN

Query Match 42.2%; Score 515.6; DB 5; Length 548;
Best Local Similarity 98.2%; Pred. No. 2.3e-128;
Matches 540; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 486 atgagctctgtgaaaaacaatgtgggagagggcctaaacatccctgtgtaatggaacc 545
|||||
Db 1 ATGAGTCTGTGAAACAAATGTGGGCAGAGGGCTAAACATCCCTGTGTGAATGGAACC 60
|||||
QY 546 acgggagctgtgctgggacagaagcatttgacatgtactctctggagatgttatgcacct 605
|||||
Db 61 ACGGGAGCTGTGTGGGACAGAGCATTTGCATGTACTCTGGAGATGTTATGCACCTA 120
|||||
QY 606 gtgaattccttaagaataatccgggggggtgcactggtgctggtgcctctctacagcat 665
|||||
Db 121 GTGAATTCCTTAAAGAAATTCGGGGGGGTGCATGTGTGTGTGCTCTCTACGACGAT 180
|||||
QY 666 ccagggacacaaatgaacgatgaagcaggagaaactctctctgaacttggggagttcctac 725
|||||
Db 181 CCAGGGACCAAAATGAACGATGAAGCAGGAAACTCTCTGACITGGGGAGTTCCTTAC 240
|||||
QY 726 gcaaaacactggcttcgggaacagctgggtcttcataggaagcacaacacacacacacac 785
|||||
Db 241 GCAAACAACTGGGGCTTCGGGACAGCTGGGTCTTCATAGGAGCCAAAGACCTCAGGGGT 300
|||||
QY 786 aaaaagccctttgagcagttctttaaagaacagccagacacacacacacacacacacacacac 845
|||||
Db 301 AAAAGCCCTTTGAGCAGTTCTTAAAGAACAGCCAGACACACAAACAAATACGAGGGATGG 360
|||||
QY 846 ccagagctgtctgagatgagggctgcatccccccgaagccatttagggtggtctgtgac 905
|||||
Db 361 CCAGAGCTGTCTGGAGATGGAGGGTGTGATCCCCCGAAGCCATTTTAGGGTGGCTGTGGC 420
|||||
QY 906 ttctctcagccaggggctgaagaagctcctgctgaacttagagtcagagcccgccgag 965
|||||
Db 421 TCTTCTCAGCCAGGGGCTGACGAAGYTCCTGCTGCATAGGACTCANAGCCCGGCA- 480
|||||
QY 966 gggctgaagagagagagaggggtgtcgtggtggaaggtgtcgcaggtccttgcacgtg 1025
|||||
Db 480 GGCTGNAGGAGGAGGAGGAGGGGTGTGCTGTGGAAGGTGCTGCAGG-CCTTTCACGCTG 538
|||||
QY 1026 tgcgcgcct 1035
|||||
Db 539 TGTGCGCCT 548
|||||

RESULT 2
AC023346/c
LOCUS 178137 bp DNA HTG 11-FEB-2000
DEFINITION Homo sapiens clone RP11-475023, *** SEQUENCING IN PROGRESS ***, 53
unordered pieces.
ACCESSION AC023346
```

```
AC023346.1 GI:5960447
HTG; HTGS_PHASE1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178137)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 178137)
Waterston,R.H.
Direct Submission
Submitted (11-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H_NH0475023.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1212: contig of 1212 bp in length
* 1213 1230: gap of unknown length
* 1231 2338: contig of 1108 bp in length
* 2339 2356: gap of unknown length
* 2357 3517: contig of 1161 bp in length
* 3518 3535: gap of unknown length
* 3536 4850: contig of 1315 bp in length
* 4851 4868: gap of unknown length
* 4869 6528: contig of 1660 bp in length
* 6529 6547: gap of unknown length
* 6547 7839: contig of 1292 bp in length
* 7839 7856: gap of unknown length
* 7857 9445: contig of 1589 bp in length
* 9446 9463: gap of unknown length
* 9464 11266: contig of 1803 bp in length
* 11267 11284: gap of unknown length
* 11285 12732: contig of 1448 bp in length
* 12733 12750: gap of unknown length
* 12751 14136: contig of 1386 bp in length
* 14137 14154: gap of unknown length
* 14155 15188: contig of 1034 bp in length
* 15189 15206: gap of unknown length
* 15207 17203: contig of 1997 bp in length
* 17204 17221: gap of unknown length
* 17222 18766: contig of 1545 bp in length
* 18767 18784: gap of unknown length
* 18785 20759: contig of 1975 bp in length
* 20760 20777: gap of unknown length
* 20778 22676: contig of 1899 bp in length
* 22677 22694: gap of unknown length
* 22695 24767: contig of 2073 bp in length
* 24768 24785: gap of unknown length
* 24786 26285: contig of 1500 bp in length
* 26286 26303: gap of unknown length
* 26304 27755: contig of 1452 bp in length
* 27756 27773: gap of unknown length
* 27774 29754: contig of 1981 bp in length
* 29755 29772: gap of unknown length
* 29773 32098: contig of 2326 bp in length
* 32099 32116: gap of unknown length
* 32117 34385: contig of 2269 bp in length
* 34386 34403: gap of unknown length
* 34404 36992: contig of 2589 bp in length
* 36993 37010: gap of unknown length
* 37011 39118: contig of 2108 bp in length
* 39119 39136: gap of unknown length
* 39137 40831: contig of 1695 bp in length
* 40832 40849: gap of unknown length
```

* 40850	43180:	contig of 2331 bp in length	QY 852	ctgctggagatgagggctgcagtcccccaagccatttttaggggtggtgctcttcc	911
* 43181	43198:	gap of unknown length			
* 43199	45573:	contig of 2375 bp in length	Db 36310	CTGCTGGAGATGGAGGGCTGCATG-CCCCGAAAGCCATTTTAGGGTGGCTGTGGCTCTTCC	36252
* 45574	45591:	gap of unknown length			
* 45592	47919:	contig of 2328 bp in length			
* 47920	47937:	gap of unknown length	QY 912	tcaagcaggggctgaagaagctctctgcctgacttaggagtcagagcccgaggggctg	971
* 47938	50743:	contig of 2806 bp in length			
* 50744	50761:	gap of unknown length	Db 36251	TCAGCCAGGGGCTGAAGAAGCTCTCTGCTGACTTAGGAGTCAGAGCCCGCAGGGGCTG	36192
* 50762	53995:	contig of 3234 bp in length			
* 53996	54013:	gap of unknown length	QY 972	aggagagagcaggggggtgctgctggtggaaggtgctgcagggtcctctgcacgctgtgctgc	1031
* 54014	56896:	contig of 2883 bp in length			
* 56897	56914:	gap of unknown length	Db 36191	AGGAGGAGGAGCAGGGGGTGCCTGCTGGAAGGTGCTGCACGCTCTTGCACGCTGTCTGC	36132
* 56915	59407:	contig of 2493 bp in length			
* 59408	59425:	gap of unknown length	QY 1032	gcctctctctctcggaaacagaaccctccacacacacatctaccctccggaagaccagctcc	1091
* 59426	62041:	contig of 2616 bp in length			
* 62042	62059:	gap of unknown length	Db 36131	GCCTCTCTCTCGGAACAGAACCCCTCCACAGACATCTTACCCGGAGACACACCTC	36072
* 62060	65315:	contig of 3256 bp in length			
* 65316	65333:	gap of unknown length	QY 1092	agagggctctctggaaccagctgctgtgagagaaatgggtgcttctgcagggagctg	1151
* 65334	67559:	contig of 2226 bp in length			
* 67560	67577:	gap of unknown length	Db 36071	AGAGGGTCTCTTCTGGAACAGCTTCTGTGAGACAATGGGTGCTTTCGTACAGGACTG	36012
* 67578	70452:	contig of 2875 bp in length			
* 70453	70470:	gap of unknown length	QY 1152	ctgacggctggtcctcagggaagacaaactgcccagactgcccacaaataaattttatt	1211
* 70471	75267:	contig of 4797 bp in length			
* 75268	75285:	gap of unknown length	Db 36011	CTGACGGCTGCTCCTGAGGAAGGACAACTGCCAGACTTGACGCCCAATTAATTTTATT	35952
* 75286	80002:	contig of 4717 bp in length	QY 1212	tttgctggt 1220	
* 80003	80020:	gap of unknown length	Db 35951	TTTCTGCTGCT 35943	
* 80021	83253:	contig of 3233 bp in length			
* 83254	83271:	gap of unknown length			
* 83272	87054:	contig of 3783 bp in length			
* 87055	87072:	gap of unknown length			
* 87073	93056:	contig of 5984 bp in length			
* 93057	93074:	gap of unknown length			
* 93075	97948:	contig of 4874 bp in length			
* 97949	97966:	gap of unknown length			
* 97967	103110:	contig of 5144 bp in length			
* 103111	103128:	gap of unknown length			
* 103129	107942:	contig of 4814 bp in length			
* 107943	107960:	gap of unknown length			
* 107961	113263:	contig of 5303 bp in length			
* 113264	113281:	gap of unknown length			
* 113282	118605:	contig of 5324 bp in length			
* 118606	118623:	gap of unknown length			
* 118624	123585:	contig of 4962 bp in length			
* 123586	123603:	gap of unknown length			
* 123604	129526:	contig of 5923 bp in length			
* 129527	129544:	gap of unknown length			
* 129545	134942:	contig of 5398 bp in length			
* 134943	134960:	gap of unknown length			
* 134961	138534:	contig of 3574 bp in length			
* 138535	138552:	gap of unknown length			
* 138553	147398:	contig of 8846 bp in length			
* 147399	147416:	gap of unknown length			
* 147417	155917:	contig of 8501 bp in length			
* 155918	155935:	gap of unknown length			
* 155936	168242:	contig of 12307 bp in length			
* 168243	168259:	gap of unknown length			
* 168260	178137:	contig of 9878 bp in length.			
FEATURES			Location/Qualifiers		
source			1..178137		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/clone="RP11-475023"		
BASE COUNT			46731 a 40913 c 41468 g 47960 t 1065 others		
ORIGIN					
Query Match			33.1%; Score 404.2; DB 55; Length 178137;		
Best Local Similarity			97.9%; Pred. No. 2.2e-98;		
Matches 420; Conservative			0; Mismatches 8; Indels 1; Gaps 1;		
Qy 792	cccttgagcagttcttaagaacagcccgagacacaaacataacagggatgcccagag	851			
Db 36370	CTCTCTCTAGTCTTAAAGAACAGCCGACACAAACAAATACGAGGATGGCCAGG	36311			

```

ETKRLFSELSRNAKELAFRDSWVFVGAKGVQVONKSPFFQHVKNKSHSNKYEGGPEALE
MEGCIPRRSTAS*
polya_signal
BASE COUNT 301 a 521 c 503 g 281 t
ORIGIN

```

Query Match	20.3%	Score 247.8	DB 9	Length 1606
Best Local Similarity	66.2%	Prod. No.3.le=56		
Matches 357	Conservative	0	Mismatches 182	Indels 0
Gaps	0			
QY	344	ctcgcgccacgaaggagatccagggttaaaaagtacaaagtgtggcctcatcaagccctgcc	403	
Db	415	CTCGGTGACTGCAGCGCCACGGCCAGGAAGTACAAAGTGTGGCTGCCCCAGCCGTGCC	474	
QY	404	agccaaactactttgsgtttaaaatctgcagtyggggccgcacacgctcgtggccctactat	463	
Db	475	TGAGGAGCACCTTGGCGCTTCCCGCTGGTCAAGCGGGCGCCAAAGTCAATTGGGCCCAAGAT	534	
QY	464	gtgctttgaagaccgcatagtatcctgtgaaaaacaattgtgggcagagggcctaaa	523	
Db	535	CTGCCTCGAGGACAAGATGCTGATGACGCGTCAAGGACAAGTGGCGCGCGGCTGAA	594	
QY	524	catacgccttgtgaatgaaacacacgggagctgtgctgggacagaagaatttgacatgta	583	
Db	595	CATPCGCCCTGTGTGAACGGGGTCAAGCGGAGCTCATCTGAGGCCCGGCCCTTTTGACATGTG	654	
QY	584	ctctggagatgttatgcaccttagtgaattcccttaagaataatccggggggtgcactggt	643	
Db	655	GGCGGAGATGTCACGACCTGTTGAAGTTATTTCGCCCACTGCACGAAGGCAACCCCTG	714	
QY	644	gctggtggcctcctacgacgatccaggggacaaaatagaacgataaaagcagagaactctt	703	
Db	715	GTTCTGGGCATCCTTACGACGACCCAGCCACCAAGATGAATTGAAGAGACCAGAAAGCTCTT	774	
QY	704	ctctgacttggggagtctcctacgcacaaacactgggcttcgcgggacagctgggtttcct	763	
Db	775	CAGTGAGCTGGGCACGAGCAACGCCAAGGAGCTGGCCCTCCGGGACAGCTGGGTCTTGT	834	
QY	764	aggagccaaagacctcaagggttaaaacccctttgacagttcttaagaacacagcccaga	823	
Db	835	CGGGGCCAAAGGTGTGCAGAACAAAGAGCCCTTTTGACGACGACGTGAAGCAAGTAA	894	
QY	824	cacaaacaaatcagggatggccagagctgctggagatggagggctgcataccccccga	882	
Db	895	CAGCAACAAAGTACGAAGGCTGCCCGAGGCGCTGGAGATGGAAGGCTGTATCCCGGA	953	

RESULT	4
LOCUS	MM072677/c
DEFINITION	Mus musculus EF-7 mRNA, partial cds.
ACCESSION	U72677
VERSION	U72677.1 GI:1872484
KEYWORDS	.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
REFERENCE	1 (bases 1 to 465) Fu.X. and Kamps,M.P. E2a-Pbx1 induces aberrant expression of tissue-specific and developmentally regulated genes when expressed in NIH 3T3 fibroblasts
TITLE	
JOURNAL	Mol. Cell. Biol. 17 (3), 1503-1512 (1997)
MEDLINE	97184476
REFERENCE	2. (bases 1 to 465) Fu.X. and Kamps,M.P.
AUTHORS	Direct Submission
TITLE	
JOURNAL	Submitted (27-SEP-1996) Pathology, University of California, Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES	Location/Qualifiers 1..465

BASE COUNT		89	129	g	109	g	138	t
ORIGIN								
<pre> /organism="Mus musculus" /strain="BALB/c" /db_xref="taxon:10090" /cell_line="NIH3T3" /cell_type="E2a-Pbx1 induced fibroblasts" complement(117..>465) /note="similar to human 2.19 protein" /codon_start=2 /product="EF-7" /protein_id="AAB51037.1" /db_xref="GI:1872495" /translation="TQGVVKKDSFDMYSGDQLLINFLEIPDSTLVLVASYPDPTK MNDKIGLFLSNLSSVAKOLGRDSWVFVGAKDLKSKSPYEQFLKNNPETNKYDQWGE LLELEGKVPKRV" </pre>								
Query Match		18.8%; Score 229.4; DB 12; Length 465;						
Best Local Similarity		77.4%; Pred. No. 2.9e-51;						
Matches 278; Conservative		0; Mismatches 81; Indels 0; Gaps 0;						
QY	546	acggagcgtgctgctggacaaagccattgacatctactctgagatgttatgcaccta	605					
DB	464	ACTGGCCAGGTGATGAAGAAGCAGCTTTGACATATGACTCTGGAGACCCCACTCCTG	405					
QY	606	gtgaattccttaagaaattcccgggggggtgcactggtgctggtggtcctctacagcagt	665					
DB	404	CTGAACCTTCTTACCGGAATTCACAGATAGCACCGCTGCTGCTAGCTCCCTATGATGAC	345					
QY	666	ccaggagccaaatgaacgatgaagcagacaaactctctgactctggaggttcctac	725					
DB	344	CCAGGACCCAAATGAACGACAAATCAACACGCTTTCTCCACCTGGGAAGTTCCTAC	285					
QY	726	gcaaaactggtggtccgggagcgtggtgttctcatagggagcgaagcctcaggggt	785					
DB	284	GCCAGCAGCTGGGCTTCGGGACAGCTGGGCTCTTTGTAGTGTCTAAACAGCTCAAGAGT	225					
QY	786	aaagcccttgagcagctcttaaaagacagccagacacaaacatacagaggtg	845					
DB	224	AAAAGCCATATGACGAGTCTTAAAGACAAACCCAGAACAAACAATATGACGGATGG	165					
QY	846	ccagagctgtagagtgagggtgctgcatgccccgaagccatttagtggtggtctgtg	904					
DB	164	CCGGAGTTGCTAGAGCTCGAGGGCTGTGCGCACGGAAAGTAATGTAGGGTGGCCTAGG	106					
RESULT	5							
LOCUS	D87120	2475 bp	mrna	PRI	11-MAR-1998			
DEFINITION	Homo sapiens mRNA for GS3786, complete cds.							
ACCESSION	D87120							
VERSION	D87120.1	GI:1507673						
KEYWORDS	GS3786.							
SOURCE	Homo sapiens cancellous bone tissue_lib:3 end-directed library osteoblast cell_line:primary-cultured cdna to mrna, clone.lib:lambda ZAP clone:GS3786.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
AUTHORS	1 (bases 1 to 2475)							
TITLE	Ohno,I.							
JOURNAL	Direct Submission							
REFERENCE	Submitted (14-AUG-1996) to the DDBJ/EMBL/GenBank databases. Ikko Ohno, Institute for Molecular and Cellular Biology, Osaka University, Molecular Genetics; 1-3 Yamada-oka, Suita, Osaka 565, Japan (E-mail:kousaku@mcb.osaka-u.ac.jp, Tel:81-6-879-7992, Fax:81-6-877-1922)							
AUTHORS	2 (bases 1 to 2475)							
TITLE	Ohno,I., Hashimoto,J., Takaoka,K., Ochi,T., Okubo,K. and Matsubara,K.							
JOURNAL	The cloning of a cdna for novel genes expressed in human osteoblast unpublished (1996)							

FEATURES	source	Location/Qualifiers
1. .2475		/organism="Homo sapiens"
/db_xref="taxon:9606"		
/cell_line="primary-cultured"		
/cell_type="osteoblast"		
/clone="GS3786"		
/clone.lib="lambda ZAP"		
/tissue_lib="3 end-directed library"		
/tissue_type="cancellous bone"		
168. .851		
CDS		
/codon_start=1		
/product="GS3786"		
/protein_id="BAAL3251.1"		
/db_xref="GI:1507674"		
/translations="MVAAGAAKLVAVAVELLTFVISOVPEIKMDASLGNLFARSL		
DTAARSKPRRYKGLSKACPEKHFAPKMASGAANYVGPKEICLEDNYVLMSGVKNVGR		
GINVALANGTSGVLFGRDFMWGQVAPFIEFLKAQDGTIVLFGTIGDGTALNDE		
ARMLIADGISTINLIGFRDNVFCGGGIKTKSPFEQHIKNNKDKTKYEGWPEVVEN		
EGCIPQKQD"		
BASE COUNT	811 a	380 c 527 g 757 t
ORIGIN		
Query Match	17.78;	Score 216.4; DB 9; Length 2475;
Best Local Similarity	63.08;	Pred. No. 8.9e-48;
Matches 334; Conservative	0;	Mismatches 196; Indels 0; Gaps 0;
QY	375	tacaagtgggcctcacaagcctgccagcccaactactttgcgtttaaaatctgcagt 434
Db	333	TATAAGTGTGGGATCTCAAAAGCTTGCCCTGAGAAGCATTTTGTCTTTAAAAATGGCAAGT 392
QY	435	ggggcgccaaagctggtggccctactatgtgtgttgaagacgcgatgatcagctc 494
Db	393	GGAGCAGCCACGCTGGTGGGACCAAAATCTGCTGGAGAGTAATGCTTTTAATGAGTGGT 452
QY	495	gtgaaaaacaatgtgggcagagggcctaaacatgcgcctggtgaatggaaccagggagct 554
Db	453	GTTAAGAAATAATGTGGGAAGAGGATCAATGTTGCCCTTGGCAATGGAAGAAACAGAGAA 512
QY	555	gtgctgggacagaagccatttgacatgctctggagatgttatgacctagtgaattc 614
Db	513	GTATTAGACACTAAATATTTTGACATGTGGGAGGAGATGTGGCACCATTTATTGAGTTT 572
QY	615	cttaagaataatccgggggtgactggtgctggtggcctcctacagacatccagggacc 674
Db	573	CTGAAGGCCCATACAAGATGGAACAATAAGTTTATGGGAACATACGATGATGGACAACC 632
QY	675	aaatgaacgatgaagcagcagaactctctgacttgggagttcctacgcacaaacaa 734
Db	633	AAACTCAATGATGAGGCACGCGCGCTCATTCGTGATTGGGGAGCACATCTATTACTAAT 692
QY	735	ctgggtctccgggacagctgggtttctcattagggagcgaagacctcaggggtaaaagcccc 794
Db	693	CTGTGTTTATAGACACAACCTGGGTCTCTGTGTGGGAAGGGCATTAAGACAAAAGCCCT 752
QY	795	tttgagcagctcttaagaacacagcccagacacacaaataacagggatggccagagctg 854
Db	753	TTTGAACACACATAAAGACAATAAAGATACAAACAATATCAAGGATGGCCTGAAGTT 812
QY	855	ctggagatggagggtgctgactgccccgaagccatttaggggtggtcgtgg 904
Db	813	GTAGAATAATGGAAGGATGCATCCCCCAAGACGACTAATGGAAATGTGG 862
RESULT	6	
HS6PDPGEN		
LOCUS	HS6PDPGEN	52173 bp DNA PRI 20-AUG-1997
DEFINITION	H.sapiens G6PD gene for glucose-6-phosphate dehydrogenase.	
ACCESSION	X55448.2	
VERSION	X55448.1	GI:450527
KEYWORDS	2-19 protein; G6PD gene; glucose-6-phosphate dehydrogenase.	
SOURCE	human.	

```

exon      13327..13364
/gene="G6PD"
/citation=[2]
number=3
13460..13568
/gene="G6PD"
/citation=[2]
number=4
14118..14335
/gene="G6PD"
/citation=[2]
number=5
15007..15165
/gene="G6PD"
number=6
15343..15468
/gene="G6PD"
number=7
15834..15927
/gene="G6PD"
number=8
16090
/gene="G6PD"
/notes="in the middle of intron 8"
/citation=[1]
replace="t"
16375..16561
/gene="G6PD"
number=9
16701..16936
/gene="G6PD"
number=10
17041..17117
/gene="G6PD"
number=11
17210
/gene="G6PD"
/notes="near the end of intron 11"
/citation=[1]
replace="c"
17223..17315
/gene="G6PD"
number=12
17413..18112
/gene="G6PD"
number=13
17860
/gene="G6PD"
/notes="exon13, 3'UTR"
/citation=[1]
replace="g"
18094..18099
/gene="G6PD"
18112
/gene="G6PD"
/citation=[2]
/evidence=experimental
19058
/notes="downstream G6PD intergenic region"
/citation=[1]
replace="t"
33000..33289
/gene="2-19 gene"
/citation=[1]
label=exon1
number=1
33000..42862
/gene="2-19"
/evidence=experimental
33000..33289
/gene="2-19 gene"
33000..42862

slg_peptide
CDS
/gene="2-19"
join(33277..33289,36113..36226,37166..37189,40439..40562,
40689..40747,41175..41225,41546..41630,41707..41833,
42131..42226)
/gene="2-19"
/citation=[1]
/codon_start=1
/label=protein2-19
/evidence=experimental
/product="2-19 protein"
/protein_id="CAA39090.1"
/db_xref="GI:452270"
/db_xref="SWISS-PROT:P98173"
/translation="MRLAGPLRIYVLVSVGVTVIIVSVTLGGPGSGFPRIQLETS
ESSVTAAPRARKYKGLPQPCPEHLAFRVYSGAANVIGPKICLEDKMLMSYKDNVG
RGLNTALVNGVSGELIEARAFDMWAGVDNLLKFIPLHEGTLVVFVSYDDPATKME
ETKLFSELGSRNAKELAFRDSWVFGAKGVQNKSPFEOHVKNKSHKYECPLEALE
MEGCIPRRSTAS"
36113..36226
/gene="2-19"
/citation=[1]
/label=exon2
number=2
37166..37189
/gene="2-19"
/citation=[1]
/label=exon3
number=3
40439..40562
/gene="2-19"
/citation=[1]
/label=exon4
number=4
40689..40747
/gene="2-19"
/citation=[1]
/label=exon5
number=5
41175..41225
/gene="2-19"
/citation=[1]
/label=exon6
number=6
41546..41630
/gene="2-19"
/citation=[1]

Query Match      5.7%; Score 70; DB 9; Length 52173;
Best Local Similarity 68.3%; Pred. No. 2.3e-08;
Matches 97; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 666 ccagggaccacaaatgaacatgaagcaggaaactctctctgacttggggagttctac 725
||| | || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41696 CCATGCTGCAGATGAATGAAGAGACCAGAAAGCTCTTCAGTGAGCTGGCAGCGAAC 41755
QY 726 gcaaacacactggctccgggacagctgggtcttcattagagccaaagacctcagggt 785
||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41756 GCCAAGGAGCTGCCCTCCGGGACAGCTGGGTGTTTGTGCGGGCCCAAGGGTGTGCAGAAC 41815
QY 786 aaagacccctttgacagtct 807
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41816 AAGAGCCCTTTGAGCAGGTAT 41837

RESULT 7
HUMFLNG6PD/c
LOCUS HUMFLNG6PD 219447 bp DNA PRI 17-MAY-1999
DEFINITION Homo sapiens chromosome X region from filament (FLN) gene to
glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds's.
ACCESSION L44140
VERSION L44140.1 GI:1203968

```

KEYWORDS 1A gene; 2.19 gene; ABP-280 gene; DNILL gene; DNase I-like protein gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene; XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin; emery-dreyfuss syndrome; filamin; glucose-6-phosphate dehydrogenase.

SOURCE ORGANISM Homo sapiens

REFERENCE AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 219447) Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N., Zuo,L., Heiner,C., Burrough,F.W., Ripetto,M., Schlössinger,D. and D'Urso,M.

TITLE Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci

JOURNAL MEDLINE Hum. Mol. Genet. 5 (5), 659-668 (1996)

COMMENT 96311563 Submitted by: Ellison Chen, Department of Genetic Technology, Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln Center Drive, Foster City, CA 94404 USA and David Schlessinger, Department of Molecular Microbiology and Center for Genetics in Medicine Washington University School of Medicine, St. Louis MO 63110 USA e-mail: ellison@genseq.aplbbio.com and david@genetics.wustl.edu Note: Gene predictions were accomplished with runs of Grail versions 1.1 and 1.2, coupled with fasta and blastx comparisons to genbank & non-redundant peptide libraries. Repeat analysis was accomplished via a censor.

FEATURES

source	Location/Qualifiers
	1..219447
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="X"
repeat_unit	4..223
	/rpt_family="Alu-Sp or Alu-Sq"
	/evidence=experimental
repeat_unit	226..503
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	641..931
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	1023..1313
	/rpt_family="Alu-Sx"
	/evidence=experimental
repeat_unit	complement(1446..1536)
	/rpt_family="Alu-J"
	/evidence=experimental
repeat_region	3095..3426
	/rpt_family="CpG Island"
	/evidence=experimental
repeat_unit	complement(3663..3952)
	/rpt_family="Alu-Sb0 or Alu-Sb1"
	/evidence=experimental
repeat_unit	4912..5046
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	5047..5333
	/rpt_family="Alu-Sb0 or Alu-Sb1"
	/evidence=experimental
repeat_unit	5346..5628

repeat_unit	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
	5632..5805
	/rpt_family="Alu-Sq"
	/evidence=experimental
repeat_unit	6363..17698
	/rpt_family="11Kb repeat1"
	/evidence=experimental
	6369..6633
	/rpt_family="Alu-Sx"
	/evidence=experimental
repeat_unit	complement(7121..7409)
	/rpt_family="Alu-Sp"
	/evidence=experimental
repeat_unit	complement(7358..7432)
	/rpt_family="Alu-Sb0 or Alu-Sb1"
	/evidence=experimental
repeat_unit	complement(7426..7701)
	/rpt_family="Alu-J"
	/evidence=experimental
	8306..8592
	/rpt_family="Alu-Sx"
	/evidence=experimental
repeat_unit	complement(8863..9029)
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	complement(9031..9317)
	/rpt_family="Alu-Sq"
	/evidence=experimental
repeat_unit	complement(9318..9442)
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	9486..9773
	/rpt_family="Alu-J"
	/evidence=experimental
repeat_unit	9996..10132
	/rpt_family="Alu-Sp"
	/evidence=experimental
	10159..10255
	/rpt_family="Alu-Sq"
	/evidence=experimental
repeat_unit	10271..10367
	/rpt_family="Alu-J"
	/evidence=experimental
repeat_unit	complement(10455..10739)
	/rpt_family="Alu-Sx"
	/evidence=experimental
repeat_unit	complement(10740..10860)
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	complement(10866..11156)
	/rpt_family="Alu-Sx"
	/evidence=experimental
repeat_unit	11223..11501
	/rpt_family="Alu-J"
	/evidence=experimental
	11638..13564
	/rpt_family="CpG Island"
	/evidence=experimental
repeat_unit	complement(11884..12163)
	/rpt_family="Alu-Sb0 or Alu-Sb1"
	/evidence=experimental
repeat_unit	complement(12195..12482)
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	complement(12668..12945)
	/rpt_family="Alu-J or Alu-S"
	/evidence=experimental
repeat_unit	complement(18360..18648)
	/rpt_family="Alu-Sb0"
	/evidence=experimental
exon	complement(18984..19488)
	/gene="FLN"

RESULT	9
I66494/c	
LOCUS	
DEFINITION	
ACCESSION	

```
VERSION      I66494.1  GI:2724471
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 7218)
AUTHORS      Dorner,F., Scheflinger,F. and Falkner,F.Gunter.
TITLE       Recombinant fowlpox virus
JOURNAL      Patent: US 5670367-A 14 23-SEP-1997;
FEATURES     Location/Qualifiers
             1..7218
             /organism="unknown"
BASE COUNT   1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match      3.8%; Score 46.6; DB 5; Length 7218;
Best Local Similarity 6.0%; Pred. No. 0.048;
Matches 25; Conservative 213; Mismatches 177; Indels 0; Gaps 0;

QY 480 atgatacagtcgtctgtgaaacaaatgtggcgagagcctaaacatcgccctgggtgaat 539
| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1447 AAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 540 gaaacacggagcgtgctgtggacagagcattgacatgtactctgagatgttatg 599
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1387 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1328
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 600 caccatagtgaattccttaagaataatccgggggggtgcactgtgctggcctctac 659
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1327 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1268
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 660 gacgaccagggaccacaaatgaacgatgaagcaggaaactctctgacttgaggagt 719
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1267 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1208
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 720 tctacgcaaaacactggcttcogggacagctgggtcttcataaggagcacaagacctc 779
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 780 aggggtaaaagcccttgacagctcttaagaacagccagccagacacaaacaacag 839
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 840 ggatggccagagcgtgctggagaggtgctgacgcccgcgaagccattttagg 894
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1087 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
AC016137/c
LOCUS
DEFINITION Homo sapiens clone RP11-90809, *** SEQUENCING IN PROGRESS ***, 30
unordered pieces.
ACCESSION AC016137
VERSION AC016137.2 GI:7007688
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 35052)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesha,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 35052)
Worley,K.C.
Direct Submission
Submitted (23-NOV-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:6466486.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: HMKR
Center clone name: RP11-90809
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 12430 bases at least Q40
Consensus quality: 19864 bases at least Q30
Consensus quality: 23509 bases at least Q20
Estimated insert size: 35052; agarose-fp estimation
Estimated insert size: 24454; sum-of-contigs estimation
Quality coverage: 0.3x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 784: contig of 784 bp in length
* 785 804: gap of unknown length
* 805 1598: contig of 794 bp in length
* 1599 1618: gap of unknown length
* 1619 2424: contig of 806 bp in length
* 2425 2444: gap of unknown length
* 2445 3277: contig of 833 bp in length
* 3278 3297: gap of unknown length
* 3298 4099: contig of 802 bp in length
* 4100 4119: gap of unknown length
* 4120 4917: contig of 798 bp in length
* 4918 4937: gap of unknown length
* 4938 5605: contig of 668 bp in length
* 5606 6721: contig of 1096 bp in length
* 6722 6741: gap of unknown length
* 6742 7433: contig of 692 bp in length
* 7434 7453: gap of unknown length
* 7454 8674: contig of 1221 bp in length
* 8675 8694: gap of unknown length
* 8695 9946: contig of 1252 bp in length
* 9947 9966: gap of unknown length
* 9967 11280: contig of 1314 bp in length
* 11281 11300: gap of unknown length
* 11301 12103: contig of 803 bp in length
* 12104 12123: gap of unknown length
* 12124 13186: contig of 1063 bp in length
* 13187 13208: gap of unknown length
```

Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabbah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A., and
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D., and
Gibbs,R.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: HMKR
Center clone name: RP11-90809
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 12430 bases at least Q40
Consensus quality: 19864 bases at least Q30
Consensus quality: 23509 bases at least Q20
Estimated insert size: 35052; agarose-fp estimation
Estimated insert size: 24454; sum-of-contigs estimation
Quality coverage: 0.3x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
* 13207 14508: contig of 1302 bp in length
* 14509 14528: gap of unknown length
* 14529 15393: contig of 865 bp in length
* 15394 15413: gap of unknown length
* 15414 16603: contig of 1190 bp in length
* 16604 16623: gap of unknown length
* 16624 17853: contig of 1230 bp in length
* 17854 17873: gap of unknown length
* 17874 19035: contig of 1162 bp in length
* 19036 19055: gap of unknown length
* 19056 19920: contig of 865 bp in length
* 19921 19940: gap of unknown length
* 19941 20420: contig of 480 bp in length
* 20421 20440: gap of unknown length
* 20441 21251: contig of 811 bp in length
* 21252 21271: gap of unknown length
* 21272 22415: contig of 1144 bp in length
* 22416 22436: gap of unknown length
* 22437 23901: contig of 1466 bp in length
* 23902 23921: gap of unknown length
* 23922 25452: contig of 1531 bp in length
* 25453 25472: gap of unknown length
* 25473 26950: contig of 1478 bp in length
* 26951 26970: gap of unknown length
* 26971 29346: contig of 2376 bp in length
* 29347 29366: gap of unknown length
* 29367 31505: contig of 2139 bp in length
* 31506 31525: gap of unknown length
* 31526 32876: contig of 1351 bp in length
* 32877 32896: gap of unknown length
* 32897 35052: contig of 2156 bp in length.
FEATURES
    source
    Location/Qualifiers
        1..35052
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-90809"
BASE COUNT      8378 a  9047 c  7874 g  9070 t   683 others
ORIGIN
Query Match      3.5%; Score 43.4; DB 57; Length 35052;
Best Local Similarity 79.0%; Pred. No. 0.34;
Matches 64; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
Qy 461 tatgtcttgaagaccgatcatcgtctgtgaaacacatgtggcagagccct 520
Db 16366 TTTCTTTCATGCTCTTCTGGTTCATGATCTCTGTGAACACATGTGGCAGAGGCT 16307
Qy 521 aaacatgcgcctggatgag 541
Db 16306 TTT-ATCGCCCTGGTGAATGG 16287
RESULT 11
AF090946 AF090946 1302 bp mRNA PRI 12-JAN-2000
LOCUS Homo sapiens clone H00688.
DEFINITION AF090946
ACCESSION AF090946
VERSION AF090946.1 GI:6690254
KEYWORDS FLI_CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1302)
Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,
Zhou,S., Liu,M. and He,F.
Functional prediction of the coding sequences of 50 new genes
deduced by analysis of cDNA clones from human fetal liver
Unpublished
JOURNAL 2 (bases 1 to 1302)
REFERENCE Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,
AUTHORS Zhou,S., Liu,M. and He,F.
```

```
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1998) Dept. of Experimental Hematology, Beijing
Institute of Radiation Medicine, 27 Taiping RD, Beijing 100850,
P.R.China
FEATURES
    source
    Location/Qualifiers
        1..1302
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="H00688"
            /tissue_type="liver"
            /dev_stage="fetus"
BASE COUNT      324 a  217 c  179 g  415 t  167 others
ORIGIN
Query Match      3.5%; Score 43; DB 51; Length 1302;
Best Local Similarity 12.0%; Pred. No. 0.45;
Matches 28; Conservative 115; Mismatches 90; Indels 0; Gaps 0;
Qy 538 atgaaccacgggagctgtgtggacagaagcatttgacatgtactctgagatgta 597
Db 545 ACKSMWTTTMYRRANCYCYCYTWTAAAWYTYYSKWWTYMWYTYMRSCMWANKR 604
Qy 598 tgcacctagtgaattctctaaagaaattccggggggtgcactggtggtggcctct 657
Db 605 AWTCCTTTTMAAAWYKKKKRAWYTTTCYKSMRSCYKRRGRSCYTWRYCYCTTTC 664
Qy 658 acgacgatccaggaccacaaatgaacgatgaagacgagaaactctctctgactgggga 717
Db 665 YKRWKGKCYMKMKSMWMTYMWYKGGCYKSMRKKRAAAARKWWTYTWKTKGRKK 724
Qy 718 gttctacgcgaacaaactgggttcgcgacagctgggtcttcacagagcc 770
Db 725 SYKSAAMCCMRSCYTAWKRWYTTTAKGSAARSMYTTTKKKKIMMKTYSSI 777
RESULT 12
HSJ963E22/c
LOCUS HSJ963E22 173823 bp DNA HTG 03-FEB-2000
DEFINITION Homo sapiens chromosome 20 clone RP5-963E22, *** SEQUENCING IN
PROGRESS ***, 16 unordered pieces.
ACCESSION AL096828
VERSION AL096828.18 GI:6982696
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173823)
Lloyd.D.
Direct Submission
Submitted (03-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced gi:6911686.
IMPORTANT: this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Contig_ID: 00572 Length: 1314bp
Contig_ID: 01063 Length: 9948bp
Contig_ID: 01189 Length: 50802bp
Contig_ID: 01224 Length: 1175bp
Contig_ID: 02145 Length: 58359bp
Contig_ID: 02238 Length: 1505bp
Contig_ID: 02506 Length: 13105bp
Contig_ID: 02914 Length: 1832bp
Contig_ID: 04122 Length: 4757bp
Contig_ID: 04133 Length: 1039bp
Contig_ID: 04135 Length: 1189bp
Contig_ID: 04137 Length: 1203bp
```


